ENS 491-492 – Graduation Project

Final Report

Project Title:

Anomaly Detection in Directed Energy Deposition (DED) Processes Using Thermal Images

Group #121 Members:

Berke Ayyıldızlı - 31018

Beyza Balota - 31232

Kerem Tatari - 29208

Supervisor(s):

Mustafa Ünel

Shawqi Mohammed Farea

Date:

18.05.2025

1. EXECUTIVE SUMMARY **2**
2. PROBLEM STATEMENT **3**
2.1. Objectives/Tasks 3
2.2. Realistic Constraints 4
3. METHODOLOGY **5**
3.1. Preprocessing 6
3.2. Supervised Models 6
a. Random Forest 6
b. XGBoost 7
c. CNN 8
3.3. Semi-Supervised Models 9
a. One Class SVM 9
b. Isolation Forest 10
3.4. Unsupervised Models 11
a. Autoencoder 11
b. DBSCAN 12
3.5. Engineering Standards 13
4. RESULTS & DISCUSSION **14**
4.1. Dataset 14
4.2. Quantitative Results 14
4.3. Key Observations 18
4.4. Trade-offs and Interpretability 20
5. IMPACT **21**
6. ETHICAL ISSUES **23**
7. PROJECT MANAGEMENT **24**
8. CONCLUSION AND FUTURE WORK **25**
9. REFERENCES **27**

1. EXECUTIVE SUMMARY

This project addresses the problem of detecting porosity-related defects in Directed Energy Deposition (DED) processes using thermal imaging. Defects such as porosity or lack of fusion significantly compromise the mechanical integrity of printed metal components. Traditional inspection methods are either too slow or invasive.

To tackle this challenge, we developed a machine learning–based anomaly detection framework capable of processing thermal frames and identifying defective instances. The dataset consisted of thermal images from a metal additive manufacturing process, with a high class imbalance and significant noise. We applied a comprehensive preprocessing pipeline, including missing value imputation, normalization, and feature extraction. Both statistical and geometrical features were evaluated, with statistical descriptors demonstrating superior performance.

We implemented and compared a range of models: Autoencoder and Density-based spatial clustering of applications with noise (DBSCAN) for the unsupervised, Random Forest, eXtreme Gradient Boosting (XGBoost) and Convolutional Neural Networks (CNN) for the supervised, and One-Class Support Vector Machine (One-Class SVM), Isolation Forest for the semi-supervised. Synthetic Minority Oversampling Technique (SMOTE) was used to mitigate class imbalance in supervised models.

Our findings showed that supervised models trained on statistical features, particularly Random Forest and XGBoost, achieved the highest accuracy and recall. Random Forest achieved the best overall performance, with an F1-score of 0.89 and recall of 0.97. The Autoencoder with a 95th percentile threshold provided the best unsupervised performance, while Isolation Forest and One-Class SVM demonstrated the potential of semi-supervised methods.

2. PROBLEM STATEMENT

Directed Energy Deposition (DED) is widely used in various fields such as aerospace, automotive, and biomedical engineering due to its ability to both build and repair metal parts with complex shapes. Nevertheless, the process has its challenges. DED systems can be unstable during operation and often produce internal defects such as porosity, cracks, or incomplete fusion, all of which can reduce mechanical strength and reliability [1], [2]. Inspired by these shortcomings, this project aims to develop a machine learning-based anomaly detection framework that can identify defects in DED processes using thermal images acquired in real time. Our goal is to investigate whether a combination of supervised and unsupervised models can offer more robust detection performance compared to existing static threshold or single-model systems, especially in conditions where labeled data is sparse, noisy, and imbalanced. Although literature has shown promising results using Autocoders, CNNs, or clustering algorithms for defect detection in additive manufacturing [3], [4], comparative studies integrating multiple types of models with realistic preprocessing and evaluation pipelines remain limited. This project addresses this gap by designing an interpretable and modular framework that can guide the future development of intelligent quality control systems for DED.

2.1. Objectives/Tasks

The project was structured around four core objectives. The first was to inspect, clean, and preprocess a thermal image dataset obtained from the Harvard Dataverse, which contains temperature frames of Ti–6Al–4V thin-walled structures fabricated via laser engineered net shaping. Preprocessing included header removal, NaN imputation, normalization, and SMOTE-based oversampling to address class imbalance. The second objective was the extraction of both statistical and geometric features from each image. In addition to metrics such as maximum, minimum, standard deviation, and interquartile range, we extracted shape-based descriptors like blob area, eccentricity, solidity, and orientation to capture structural variations in the melt pool. The third objective involved the development and evaluation of a broad range of models: unsupervised (Autoencoders, DBSCAN), supervised (Random Forest, XGBoost, CNNs), and semi-supervised approaches (One-Class SVM, Isolation Forest). All models were trained using stratified sampling and evaluated using standard classification metrics such as precision, recall, and F1-score. The final objective was to compare the effectiveness of these models in detecting porosity-related anomalies, identify their respective strengths and weaknesses, and recommend a detection strategy most suited for future industrial deployment.

2.2. Realistic Constraints

The project was designed with several realistic constraints in mind. Economically, the use of open-source libraries such as Scikit-learn, TensorFlow, and Keras eliminated licensing costs and ensured reproducibility. The dataset used was publicly available and ethically licensed, reducing both cost and legal risk. Since the work was entirely computational, it did not involve any environmental impact or pose health and safety risks typically associated with physical testing or hardware integration. A major technical constraint was the high class imbalance in the dataset, where only approximately 4.5% of samples represented defective frames. This was addressed through the use of SMOTE, which synthetically augments the minority class to reduce bias in supervised learning models [5]. Another challenge was ensuring that the models could generalize beyond this dataset, given that real-world DED systems vary in material, resolution, and process parameters. To mitigate this, the project prioritized statistical features that have demonstrated transferability across thermal datasets in manufacturing research [6]. Finally, the project adhered to relevant ethical and scientific standards, particularly those promoted by the IEEE for autonomous and intelligent systems used in high-risk domains, by ensuring the system was designed to support—not replace—human decision-making in quality assurance workflows [7], [8].

3. METHODOLOGY

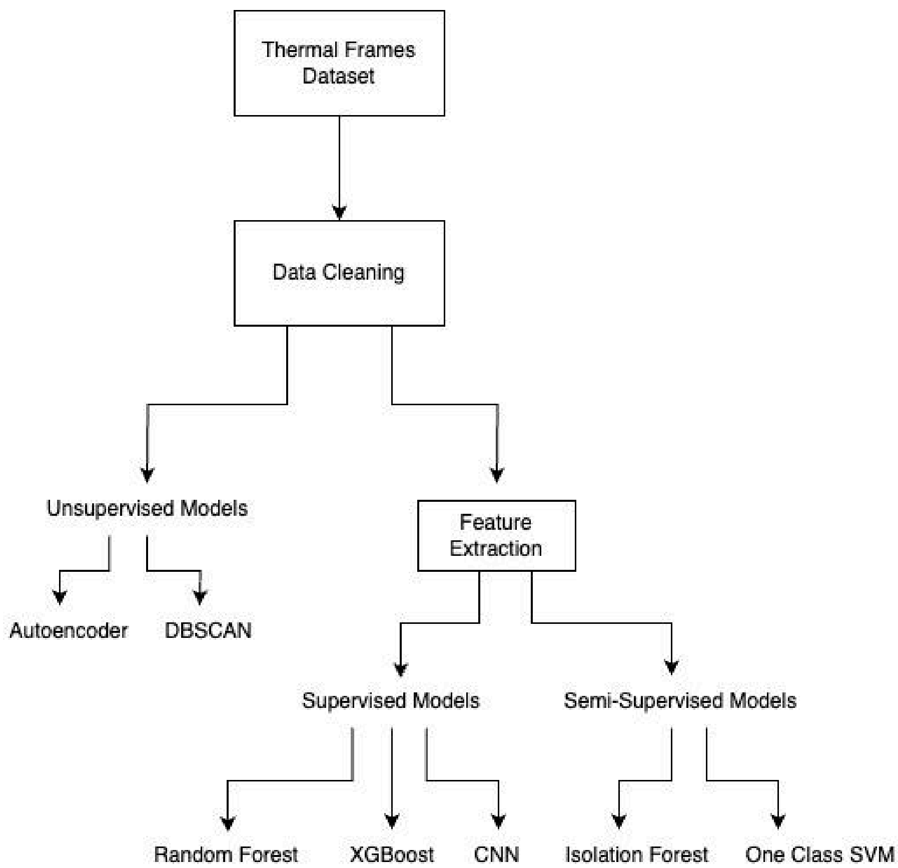

Fig. 1. Methodology Pipeline

3.1. Preprocessing

To ensure optimal performance, a robust preprocessing pipeline was developed for the raw thermal images of the DED process. The images often contained zero valued or missing pixels, making it necessary for a data cleaning phase. These missing values are all replaced with NaNs and changed to the column means using imputation strategy. After this operation, the min-max normalization was applied to scale all pixel values to the [0, 1] range. This step was particularly important for gradient based optimization user models, such as CNN and XGBoost.

To reduce high dimensionality of the thermal frames, next a feature extraction phase is implemented. As a first step, 11 statistical features were extracted, these consist of features like min, max and mean temperatures, quartiles and IQR, standard deviation, median and skewness. Later, geometrical features were also introduced to enrich the representation. These included blob-based properties such as count and area of high temperature regions, eccentricity, extent and solidity. To address the class imbalance problem SMOTE was mainly utilised during training. This helped generate synthetic positive instances to enhance model sensitivity to rare anomalies [9].

3.2. Supervised Models

a. Random Forest

A robust ensemble learning algorithm based on decision trees, Random Forest was employed as a supervised classifier to detect porosity-related anomalies from thermal frames [10]. A 5-fold stratified cross-validation strategy was used to evaluate the model's performance, ensuring class balance and full utilization of the dataset across all folds.

The model was trained using 11 statistical features extracted from each thermal frame. Although geometrical features were also explored—both independently and in combination with statistical features—models trained solely on statistical features consistently outperformed other configurations.

Additionally, experiments were conducted with top-ranked subsets of features. However, these smaller subsets failed to achieve the same level of accuracy and generalizability as the full set of statistical features. As a result, the 11 statistical features were selected as the final input representation for the Random Forest classifier.

Key model configurations and feature details are summarized in Figure 2.

Random Forest Model Configuration

n_estimators	max_depth	min_samples_split	min_samples_leaf
200	5	5	2
max_features	**class_weight**	**random_state**	
sqrt'	balanced'	42	

Fig. 2. The Random Forest Classifier

b. XGBoost

A gradient boosting framework known for its high performance and scalability on structured data, XGBoost was employed as a supervised classifier to detect porosity anomalies using statistically extracted features [11]. The model was trained and evaluated using a 5-fold stratified cross-validation strategy, mirroring the Random Forest setup to ensure a fair and consistent comparison across all data subsets.

Similar to the Random Forest experiments, geometrical features were also tested both independently and in combination with statistical features. However, statistical features alone consistently yielded better classification performance. As a result, only statistical descriptors were retained in the final input feature set for XGBoost.

In addition, different subsets of features—such as the top 5 or top 7 features based on importance scores—were also explored. Nevertheless, the full set of statistical features proved more robust and generalizable across folds, leading to its selection for the final model.

To address class imbalance, oversampling techniques were applied within each training fold to increase the representation of the minority class. The hyperparameters of the XGBoost model were fine-tuned based on cross-validation results to maximize performance. The finalized model configuration is summarized in Figure 3.

XGBoost Model Configuration

n_estimators	learning_rate	max_depth	reg_lambda
100	0.05	3	1
subsample	**colsample_by tree**	**reg_alpha**	
0.7	0.7	0.5	

Fig. 3. The XGBoost Classifier

c. CNN

Convolutional Neural Networks (CNNs) were implemented as part of the supervised learning pipeline to directly process raw thermal image frames and classify them into defective and non-defective categories. Unlike tree-based models that rely on handcrafted statistical features, CNNs can autonomously learn and extract spatial patterns relevant to defect characteristics from raw pixel data.

The model architecture includes two convolutional layers, max-pooling, dropout, and dense layers, finalized with a sigmoid output for binary classification. Key configuration details such as optimizer, loss function, and training parameters are summarized in Figure 4. The model was trained on a stratified 70–30 split and class imbalance was addressed using SMOTE during dataset preparation.

CNN_Model_Configuration

Conv Layers	Pooling	Dropout	Dense Layer	Output
2 layers (16 & 32 filters) 3×3 kernel, ReLU	MaxPooling2D	0.3	64 units, ReLU	1 unit, Sigmoid
Optimizer	**Loss**	**Epochs / Batch**	**Train/Test Split**	**Imbalance Handling**
Adam (lr=0.001)	Binary Cross-Entropy	10 / 16	70% / 30%	SMOTE

Fig. 4. Convolutional Neural Networks (CNN)

3.3. Semi-Supervised Models

a. One Class SVM

One-Class Support Vector Machine (OCSVM) was implemented as a semi-supervised approach for anomaly detection, trained exclusively on the normal (non-defective) subset of the labeled data. The key idea is to learn the boundary of the "normal" class and then apply the trained model to both labeled and unlabeled data to detect deviations from this learned representation of normality.

The input features consisted of the top 5 most important statistical descriptors extracted from thermal images: IQR, Std_Temp, Q1, Min_Temp, and Median_Temp. These were normalized using MinMax scaling. The OCSVM was configured with an RBF kernel, nu = 0.035 (representing the expected proportion of anomalies), and gamma = 0.05.

Once trained on normal samples (label = 0), the model was used to predict anomalies across the full dataset. Its outputs were processed via decision function thresholding, where predictions were binarized based on the 3.5th percentile of the decision scores to maximize F1 performance. These predictions were mapped to binary labels (normal (0) and anomaly (1)) to enable performance evaluation and integration into the larger detection framework. The trained model was also applied to unlabeled data to generate pseudo-labels, supporting potential downstream analysis. This method enables the detection of porosity-related defects even without labeled defective data and provides a foundation for semi-supervised strategies in scenarios with scarce annotations.

b. Isolation Forest

Isolation Forest was applied as a semi-supervised anomaly detection method, leveraging its efficiency in identifying outliers within high-dimensional feature spaces. The model was trained exclusively on the normal (non-defective) subset of the labeled dataset, without using label values during training, enabling it to model the distribution of the majority class.

The input consisted of the top 5 most important statistical features extracted from thermal images: IQR, Std_Temp, Q1, Min_Temp, and Median_Temp. These features were normalized using MinMax scaling prior to training. The Isolation Forest was configured with 100 estimators and a contamination parameter of 0.045, which was empirically aligned with the expected anomaly rate in the labeled data.

Once trained, the model predicted anomaly scores for the labeled dataset. Output values were thresholded at the 4.5th percentile of decision function scores to produce binary labels (0 for normal, 1 for anomaly), enabling performance evaluation on labeled samples. These pseudo-labels can be further utilized in downstream learning steps or comparative analyses with other detection strategies.

3.4. Unsupervised Models

a. Autoencoder

The autoencoder model in this study was trained in an unsupervised manner to learn the typical thermal patterns of melt pool frames. The input consisted of the flattened temperature values from frames. The architecture featured a single hidden layer with 64 encoding neurons using ReLU activation, followed by a symmetric decoding layer using sigmoid activation. The model was compiled with the Adam optimizer and mean squared error (MSE) loss function.

The autoencoder was trained on all available data for 50 epochs with a batch size of 16. After training, each thermal frame was passed through the network to generate a reconstructed version. As the deviations between the input and reconstructed output serve a basis for identification of the anomalies, the reconstruction error (MSE between the original and the reconstructed frame) was used as an anomaly score [12]. Two thresholds were defined based on the 95th and 99th percentiles of the MSE distribution. Frames exceeding these thresholds were flagged as defective.

In addition to the standard fully connected autoencoder, we also experimented with a convolutional autoencoder architecture, which is commonly used in image-based anomaly detection tasks. However, in our case, the traditional fully connected autoencoder outperformed the convolutional variant in terms of anomaly detection accuracy. As a result, we proceeded with the traditional autoencoder for the final evaluation.

b. DBSCAN

DBSCAN was employed as an unsupervised anomaly detection technique to identify defective thermal frames in the DED process. Unlike the previous approach that operated on flattened thermal images with dimensionality reduction via PCA, the updated method leverages statistical features extracted from each frame to improve performance and interpretability. Specifically, the top seven most informative features were selected based on prior analysis: IQR, Std_Temp, Q1, Min_Temp, Median_Temp, and Skewness.

Before clustering, missing values in the feature set were handled using mean imputation, and the feature space was standardized using z-score normalization. To identify the optimal epsilon parameter for DBSCAN, a sweep over a range of values (0.1 to 1.0) was performed. For each eps value, DBSCAN was applied to the preprocessed feature space, and frames labeled as noise points were classified as anomalies. The F1-score for detecting true porosity defects was computed at each step to evaluate performance.

The optimal eps was selected based on the highest F1-score obtained during the sweep. Using this best-performing configuration, DBSCAN was rerun to generate final predictions, which were then evaluated using standard classification metrics. Compared to the initial image-based implementation, this feature-based DBSCAN approach demonstrated improved anomaly detection accuracy, highlighting the effectiveness of leveraging descriptive thermal statistics over raw pixel intensities.

3.5. Engineering Standards

The development of this anomaly detection framework aligns with several key engineering standards that support the quality, consistency, and traceability of additive manufacturing systems. Specifically, ASTM F2924 provides comprehensive guidelines for the processing and evaluation of titanium and titanium alloy components produced via additive manufacturing. By adhering to this standard, the study ensures that defect detection mechanisms are grounded in accepted material behavior and testing protocols, contributing to the standardization of DED-based manufacturing processes [13].

To maintain consistency and interoperability across different stages of data collection and analysis, ISO 10303-242 (STEP AP 242) was also considered. This standard facilitates the structured exchange of product and process data, enabling seamless integration of thermal frame features and metadata across monitoring and diagnostic systems [14]. Additionally, the evaluation of model performance was informed by ISO 5725-1, which outlines principles for assessing measurement precision and trueness. These guidelines supported the rigorous benchmarking of the machine learning models, especially when comparing predictions across multiple folds and model types [15].

Finally, since data integrity is a critical concern in industrial environments, ISO/IEC 27001 was taken into account to ensure best practices for managing the confidentiality and security of the thermal image datasets. While not directly applied within the model training pipeline, this standard provides a framework for safely handling sensitive manufacturing data and securing the infrastructure that supports defect monitoring workflows [16].

4. RESULTS & DISCUSSION

4.1. Dataset

The dataset used in this study consists of thermal images capturing the Directed Energy Deposition (DED) process of Ti–6Al–4V thin-walled structures, sourced from the publicly available Harvard Dataverse. Each image corresponds to a temperature frame and is labeled based on porosity presence, with only ~4.5% of samples representing defective cases, resulting in a highly imbalanced distribution. A comprehensive preprocessing pipeline was applied, including NaN imputation, normalization, and the extraction of statistical and geometric features to enable effective model training and evaluation.

4.2. Quantitative Results

To evaluate the effectiveness of each modeling approach, performance metrics including precision, recall, F1-score, and accuracy were computed across supervised, semi-supervised, and unsupervised models. The results reflect the models' ability to detect porosity-related defects in thermal frames using either raw image inputs (CNN) or extracted statistical descriptors.

Confusion matrices summarizing prediction outcomes are presented in Figure 5, Figure 6, Figure 7, grouped by learning category (supervised, semi-supervised, unsupervised). These matrices provide insights into the trade-offs between true positive detection and false alarms for porosity detection.

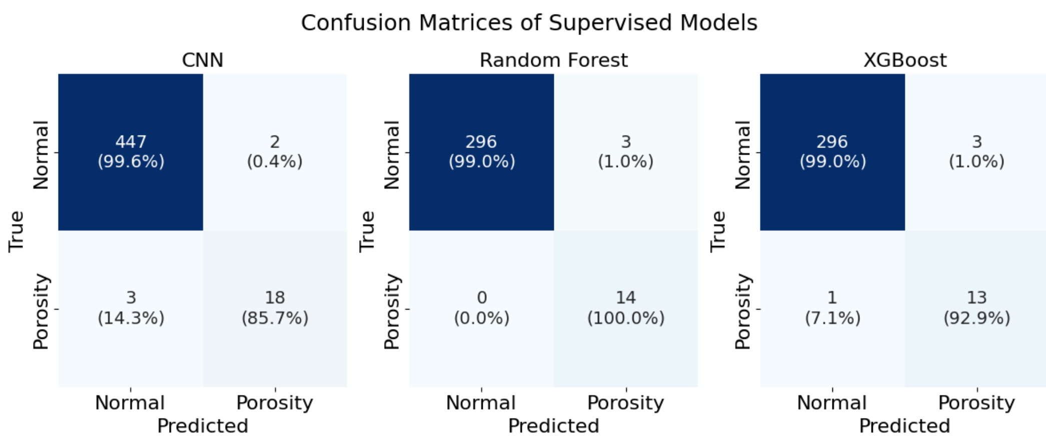

Fig. 5. Confusion Matrices of Supervised Models

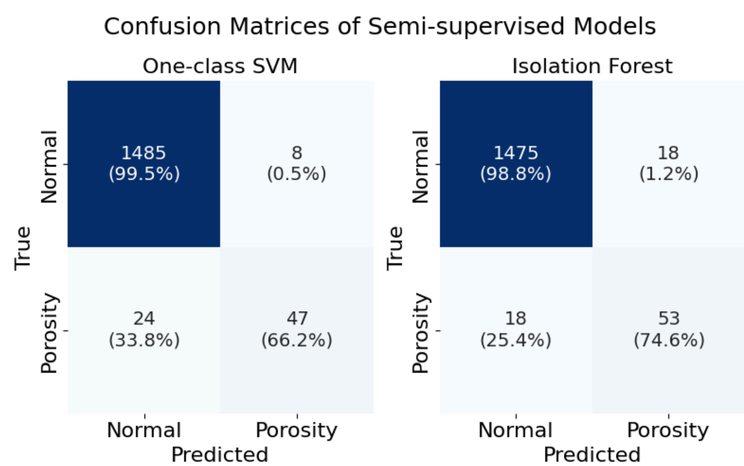

Fig. 6. Confusion Matrices of Semi-supervised Models

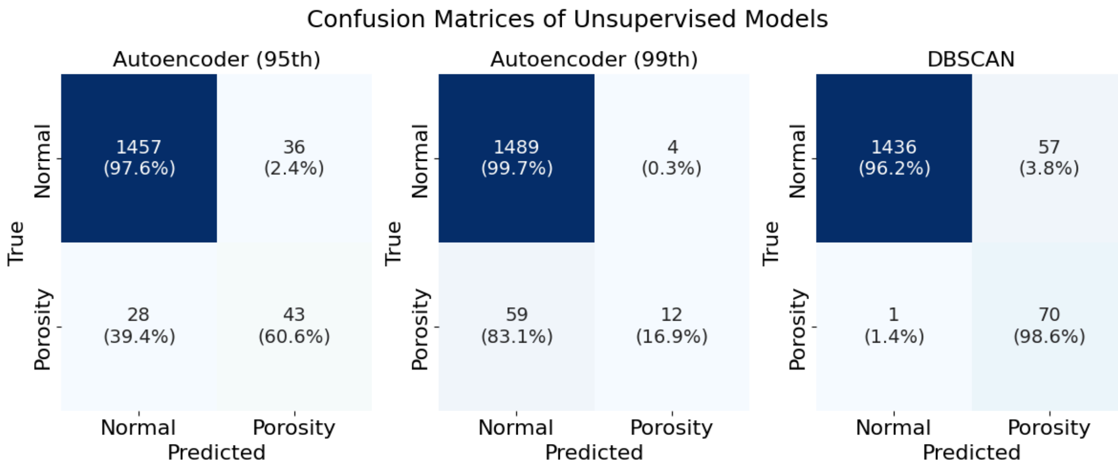

Fig. 7. Confusion Matrices of Unsupervised Models

Tables I–III detail the numerical performance metrics of each model. Among supervised models, CNN and XGBoost delivered strong and balanced results, benefiting from optimized configurations. Random Forest achieved the highest recall (0.97), suggesting strong sensitivity to porosity, though this may come at the cost of precision and potential overfitting due to limited positive samples.

In the semi-supervised category, One-Class SVM achieved higher precision (0.85) but had lower recall (0.66), indicating a more conservative detection strategy that minimizes false positives but may miss some defects. In contrast, Isolation Forest reached a balanced recall and precision of 0.75, resulting in the same F1-score (0.75) as One-Class SVM. These results highlight the trade-off between capturing more true positives and reducing false alarms, with Isolation Forest providing a more evenly balanced performance.

In the unsupervised category, Autoencoder (95th percentile) outperformed the 99th percentile setting and DBSCAN, achieving a more balanced trade-off. DBSCAN showed improved recall (0.99) and F1-score (0.71) following recent parameter tuning, making it the strongest among the unsupervised variants.

Table I

Supervised Models Results

Model	Precision	Recall	F1-Score	Accuracy
Random Forest (cross-validated)	0.83	0.97	0.89	0.99
XGBOOST (cross-validated)	0.84	0.92	0.87	0.99
CNN	0.90	0.86	0.88	0.99

Table II

Semi-supervised Models Results

Model	Precision	Recall	F1-Score	Accuracy
One-Class SVM	0.85	0.66	0.75	0.98
Isolation Forest	0.75	0.75	0.75	0.98

Table III

Unsupervised Models Results

Model	Precision	Recall	F1-Score	Accuracy
Autoencoder (99th)	0.75	0.17	0.28	0.96
Autoencoder (95th)	0.54	0.61	0.57	0.96
DBSCAN	0.55	0.99	0.71	0.96

The models also revealed varying sensitivities to the highly imbalanced nature of the data. While SMOTE-based oversampling helped improve minority class detection in supervised training, semi-supervised and unsupervised methods operated without label-aware balancing and relied solely on underlying data distributions.

4.3. Key Observations

Among supervised models, CNN and XGBoost delivered strong results, supported by SMOTE-based balancing and optimized configurations. Random Forest achieved the highest F1 score of all, and perfect recall on the minority class, though this may reflect overfitting due to limited test samples.

In the semi-supervised group, One-Class SVM achieved higher precision (0.85) but lower recall (0.66), indicating a tendency to avoid false positives. In contrast, Isolation Forest yielded a more balanced trade-off, achieving both precision and recall of 0.75, resulting in an identical F1-score (0.75). This demonstrates Isolation Forest's ability to detect a comparable number of true positives while maintaining fewer false alarms.

For unsupervised models, the Autoencoder (95th percentile) performed better than its stricter 99th percentile variant, indicating that overly conservative thresholds may under detect anomalies.. DBSCAN's performance improved significantly, achieving high recall (0.99) and a reasonable F1-score (0.71), although it remains sensitive to parameter tuning and input dimensionality.

Figure 8 shows feature importance plots for Random Forest and XGBoost, identifying IQR, standard deviation, and Q1 as key descriptors. The correlation heatmap in Figure 9 further illustrates inter-feature relationships and their link to porosity, supporting interpretability and guiding future feature selection.

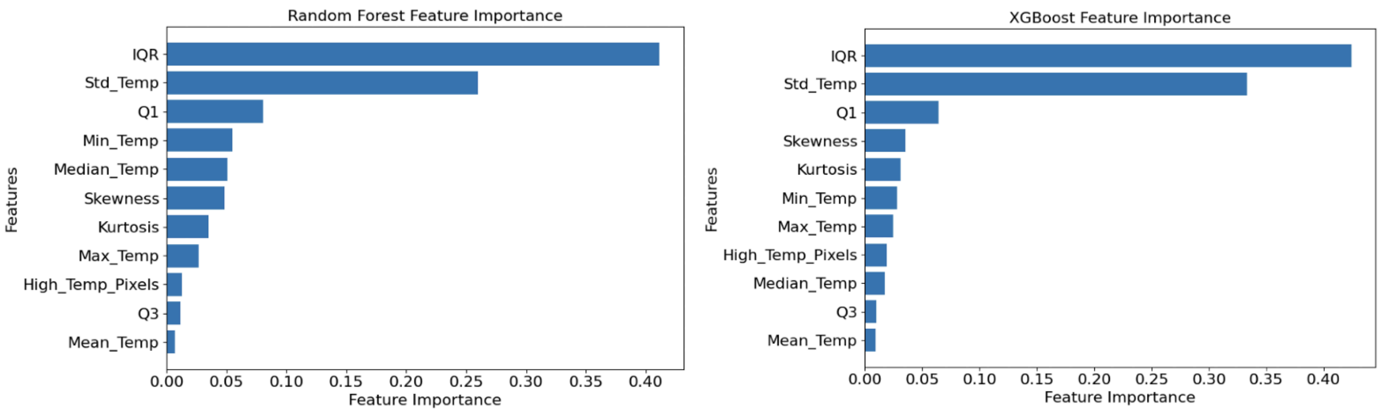

Fig. 8. Feature Importance Plots for Random Forest vs. XGBoost

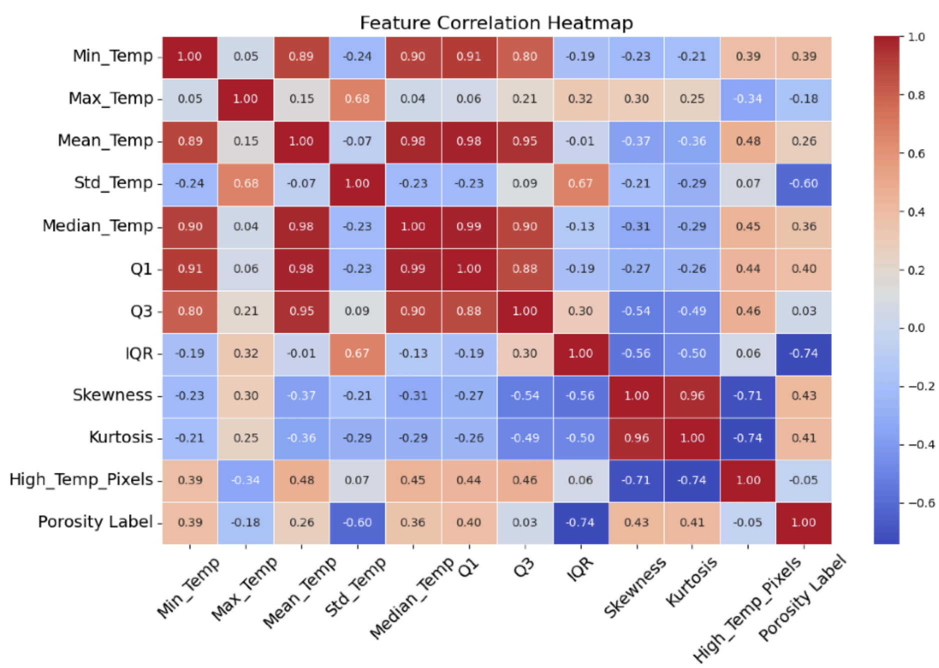

Fig. 9. Feature Correlation Heatmap

In summary, supervised models are ideal when labeled data is available, while semi-supervised methods offer a strong alternative in low-label contexts, and unsupervised models remain viable when annotations are unavailable.

4.4. Trade-offs and Interpretability

Each modeling strategy in this study presents distinct trade-offs in terms of accuracy, recall, data dependency, and interpretability. Supervised models achieved the highest overall performance (F1-scores ~0.87–0.89), benefiting from labeled data, optimized configurations, and SMOTE-based balancing to address class imbalance. However, this came at the cost of increased annotation effort, reduced flexibility, and potential overfitting risks, especially with limited positive samples.

Semi-supervised models, including One-Class SVM and Isolation Forest, provided a middle ground by leveraging unlabeled data while only requiring minimal label guidance. These models successfully identified anomalies even in the absence of defective labels, offering a practical alternative for early-stage quality monitoring where annotations are scarce. While One-Class SVM favored high precision (0.85) at the expense of recall (0.66), Isolation Forest offered a more balanced profile with both precision and recall at 0.75, resulting in equal F1-scores.

Unsupervised approaches such as DBSCAN and Autoencoder offered greater adaptability without requiring labels but showed sensitivity to thresholds and parameters. The Autoencoder's performance, for instance, varied significantly between the 95th and 99th percentile thresholds, reflecting a trade-off between recall and false-positive rates. DBSCAN initially struggled with subtle or noisy defect patterns due to sensitivity to ε and min_samples. However, after parameter tuning, it achieved strong recall (0.99) and a reasonable F1-score (0.71), demonstrating its potential when carefully configured.

From an interpretability perspective, supervised tree-based models stood out by offering clear feature importance rankings—highlighting IQR, standard deviation, and Q1 as the most influential descriptors (see Figure 8). This transparency is critical in industrial contexts where traceable model decisions are essential. In contrast, unsupervised models, while more adaptable, often lacked explainability due to their reliance on latent representations or density-based heuristics.

In deployment scenarios, unsupervised and semi-supervised models may be preferable for rapid integration in dynamic environments, whereas supervised models are better suited for well-annotated, high-precision settings. A hybrid strategy—using lightweight anomaly detection methods for continuous monitoring and switching to supervised refinement as annotations accumulate—may offer the best trade-off between robustness and practicality.

5. IMPACT

This project contributes across scientific, technological, and socio-economic areas by developing a machine learning–based anomaly detection framework for Directed Energy Deposition processes. From a scientific perspective, it offers a structured comparison of both supervised and unsupervised models —including Random Forest, XGBoost, CNN, Autoencoders, and DBSCAN— on thermal imaging data that is highly imbalanced and noisy.

The results provide useful benchmarks that can support future research on real-time defect detection in manufacturing environments [17], [18].

On the technological side, the project shows that thermal image sequences —often overlooked in industrial monitoring— can be used effectively with interpretable and scalable machine learning pipelines to detect porosity-related anomalies. This helps reduce the need for expensive post-process inspections like XCT and supports the broader goal of intelligent, edge-computing–based quality control in modern production systems, as emphasized by IEEE [19].

From a socio-economic angle, applying non-contact, real-time defect detection in metal additive manufacturing could help reduce material waste, lower energy use, and minimize production downtime. It's especially relevant in safety-critical sectors such as aerospace, healthcare, and defense, where catching issues early is vital [20]. While the project did not focus on commercialization, the approach —combining preprocessing techniques (like imputation and normalization), statistical feature extraction, and SMOTE-based class balancing— can be adapted for future industrial use or academic extensions in areas like predictive maintenance and autonomous manufacturing.

All tools and datasets used in the project were open-source or publicly licensed, and no Freedom-to-Use issues were identified [21]. The combination of interpretable features with deep learning models adds to the project's potential to support both academic research and real-world industrial systems, especially in the context of Industry 4.0 and innovations in additive manufacturing [22], [23].

6. ETHICAL ISSUES

The execution of this project fully adhered to ethical standards in academic and engineering research practices. The thermal image dataset used was sourced from the publicly available Harvard Dataverse repository, specifically from the work of Zemiela et al. [24], and contained no personally identifiable or sensitive industrial information. Its use respected all applicable data privacy, intellectual property, and open-access licensing guidelines. All machine learning models developed throughout the project relied exclusively on open-source libraries such as Scikit-learn, TensorFlow, and Keras, and no proprietary algorithms or restricted-access components were incorporated into the work. Methodological processes, including data preprocessing, feature extraction, model training, and evaluation, were thoroughly documented to support reproducibility and transparency, following best practices to prevent bias and ensure objective evaluation through established metrics such as precision, recall, and F1-score [25], [26].

No physical experiments or direct interactions with manufacturing systems were involved during this study, thus eliminating health, safety, or environmental risks. However, it is acknowledged that in future industrial applications, overreliance on machine learning outputs without appropriate human oversight could introduce critical safety risks, particularly in sectors like aerospace or biomedical manufacturing. In accordance with IEEE's ethical guidance [27], the system developed in this project is intended to support, not replace, expert decision-making within broader quality assurance workflows. Furthermore, there are no known freedom-to-use issues associated with the tools or datasets employed; no patented algorithms, proprietary datasets, or commercial software were used. All components of the proposed framework are suitable for academic sharing and potential industrial adaptation under open licensing conditions.

In conclusion, no ethical conflicts or violations were identified, and the project consistently upheld the principles of openness, safety, transparency, and responsible innovation.

7. PROJECT MANAGEMENT

We originally planned to rely mostly on unsupervised learning—autoencoders, in particular—for anomaly detection on thermal images. That seemed promising at first. To stay organized, we created a Gantt chart that divided our work into several parts: preprocessing, building the models, testing them, and finally reporting. Everyone got tasks based on what they were best at and when they had time, and we used shared spreadsheets to track everything and ensure accountability throughout the development cycle.

But once we got into the actual data, it became clear things weren't going to be that straightforward. The dataset had big issues—especially class imbalance and sparsity—which limited what we could do with unsupervised models. Because of that, we added supervised models like Random Forest, XGBoost, and CNNs, and began exploring semi-supervised techniques such as One-Class SVM and Isolation Forest. We also enhanced some models with geometric features—like blob area, eccentricity, and solidity—extracted from the melt pool to improve anomaly detection. That meant adding new steps like feature extraction and balancing the dataset using SMOTE. These changes naturally affected how tasks were divided, and we had to stay in closer contact than planned, often making quick adjustments.

In retrospect, dealing with these shifts became one of the more meaningful learning experiences. It pushed us to balance changing priorities, maintain momentum, and keep communication consistent when unexpected obstacles arose. Weekly meetings and updating our task board helped a lot. In the end, the way we worked evolved into something more flexible and realistic, which helped us handle all the complications we didn't see coming.

8. CONCLUSION AND FUTURE WORK

This study presented a comprehensive machine learning framework for detecting porosity-related anomalies in Directed Energy Deposition (DED) using thermal imaging data. By leveraging unsupervised, supervised, and semi-supervised models, we demonstrated that supervised classifiers trained on statistically extracted features—particularly Random Forest and XGBoost—achieved the most reliable performance, with Random Forest reaching an average accuracy of 98.91% and an F1-score of 0.8949 for the anomaly class.

Unsupervised models like the autoencoder performed best when thresholding was applied based on reconstruction error distribution, while DBSCAN benefited significantly from operating on extracted features instead of high-dimensional raw images. Semi-supervised methods, including One-Class SVM and Isolation Forest, were enhanced through focused feature selection and ensemble strategies, yielding competitive results despite limited label supervision.

To address prior limitations, we implemented key improvements: such as retraining DBSCAN on structured features, and re-evaluating supervised models using top-ranked features. These steps enhanced both performance and model interpretability.

Despite these advancements, further fine-tuning is possible. Model sensitivity to hyperparameters, dataset variability, and domain-specific feature engineering remain important considerations. Future efforts may focus on extending the framework to different DED systems and materials, improving robustness, and advancing semi-supervised learning to better handle label-scarce scenarios in real-world manufacturing environments.

9. REFERENCES

[1] R. Tapia and H. Elwany, “A review on process monitoring and control in metal-based additive manufacturing,” *Journal of Manufacturing Science and Engineering*, vol. 136, no. 6, pp. 060801, 2014.

[2] T. DebRoy, H. L. Wei, J. S. Zuback, 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.

[3] M. Khanzadeh, S. Chowdhury, M. Marufuzzaman, M. A. Tschopp, and L. Bian, “Porosity prediction: Supervised-learning of thermal history for direct laser deposition,” *J. Manuf. Syst.*, vol. 47, pp. 69–82, 2018.

[4] K. Chen, P. Zhang, H. Yan, G. Chen, T. Sun, Q. Lu, Y. Chen, and H. Shi, “A review of machine learning in additive manufacturing: Design and process,” *Int. J. Adv. Manuf. Technol.*, vol. 135, no. 3–4, pp. 1051–1087, 2024.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[6] M. Trovato, L. Belluomo, M. Bici, M. Prist, F. Campana, and P. Cicconi, “Machine learning in design for additive manufacturing: A state-of-the-art discussion for a support tool in product design lifecycle,” *Int. J. Adv. Manuf. Technol.*, vol. 137, no. 5–6, pp. 2157–2180, Mar. 2025.

[7] R. Chatila and J. C. Havens, “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems,” in *Robotics and Well-Being*, M. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. Tokhi, and E. Kadar, Eds. Cham, Switzerland: Springer, 2019, pp. 11–16.

[8] IEEE Standard 2755-2017, *IEEE Guide for Terms and Concepts in Intelligent Process Automation*, IEEE, New York, NY, USA, 2017, pp. 1–16. [Online]. Available: https://ieeexplore.ieee.org/document/8070671

[9] B. Ayyildizli, B. Balota, K. Tatari, ”ENS492 Model Training Code," Personal Archive, 2025.

[10] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

[12] C. Zhou and R. C. Paffenroth, “Anomaly detection with robust deep autoencoders,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017, pp. 665–674.

[13]ASTM F2924, "Standard Specification for Additive Manufacturing Titanium and Titanium Alloy Components," ASTM International. Accessed: May. 15, 2025. [Online]. Available: https://www.astm.org/f2924-14r21.html

[14] ISO 10303-242, "Industrial Automation Systems and Integration – Product Data Representation and Exchange," ISO. Accessed: May. 15, 2025. [Online]. Available: https://www.iso.org/standard/84667.html

[15] ISO 5725-1, "Accuracy (Trueness and Precision) of Measurement Methods and Results," ISO. Accessed: May. 15, 2025 [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-2:v1:en

[16] ISO/IEC 27001, "Information Security Management," ISO. Accessed: May. 15, 2025 [Online]. Available: https://www.iso.org/standard/27001

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[18] M. Khanzadeh, S. Chowdhury, M. Marufuzzaman, M. A. Tschopp, and L. Bian, “Porosity prediction: Supervised-learning of thermal history for direct laser deposition,” *J. Manuf. Syst.*, vol. 47, pp. 69–82, 2018.

[19] IEEE Standard 2755-2017, *IEEE Guide for Terms and Concepts in Intelligent Process Automation*, IEEE, New York, NY, USA, 2017, pp. 1–16. [Online]. Available: https://ieeexplore.ieee.org/document/8070671

[20] R. Tapia and H. Elwany, “A review on process monitoring and control in metal-based additive manufacturing,” *Journal of Manufacturing Science and Engineering*, vol. 136, no. 6, pp. 060801, 2014.

[21] Z. Zamiela, W. Tian, S. Guo, and L. Bian, "Thermal-Porosity Characterization Data of Additively Manufactured Ti–6Al–4V Thin-walled Structure via Laser Engineered Net Shaping, "*Harvard Dataverse*, V1, 2023. [Online]. Available: https://doi.org/10.7910/DVN/BWHYEH

[22] R. Chatila and J. C. Havens, “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems,” in *Robotics and Well-Being*, M. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. Tokhi, and E. Kadar, Eds. Cham, Switzerland: Springer, 2019, pp. 11–16.

[23] K. Chen, P. Zhang, H. Yan, G. Chen, T. Sun, Q. Lu, Y. Chen, and H. Shi, “A review of machine learning in additive manufacturing: Design and process,” *Int. J. Adv. Manuf. Technol.*, vol. 135, no. 3–4, pp. 1051–1087, 2024.

[24] Z. Zamiela, W. Tian, S. Guo, and L. Bian, "Thermal-Porosity Characterization Data of Additively Manufactured Ti–6Al–4V Thin-walled Structure via Laser Engineered Net Shaping, "*Harvard Dataverse*, V1, 2023. [Online]. Available: https://doi.org/10.7910/DVN/BWHYEH

[25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[26] D. M. W. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” *arXiv preprint* arXiv:2010.16061, 2020. [Online]. Available: https://arxiv.org/abs/2010.16061

[27] R. Chatila and J. C. Havens, “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems,” in *Robotics and Well-Being*, M. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. Tokhi, and E. Kadar, Eds. Cham, Switzerland: Springer, 2019, pp. 11–16.