APPLICATION OF MACHINE LEARNING FOR DISTORTION PREDICTION IN ADDITIVE MANUFACTURING

A PROJECT REPORT

Submitted to

Amrita Vishwa Vidyapeetham

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY IN AUTOMATION & ROBOTICS ENGINEERING

By

SUDHAKAR. R

(CH.EN.U4ARE22016)

Submitted to

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF ENGINEERING

CHENNAI – 601103

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

3D printing, also known as **Additive Manufacturing (AM)**, has revolutionized the way products are designed and manufactured. Unlike traditional subtractive manufacturing methods, which remove material from a solid block, 3D printing builds objects layer by layer from digital designs. This approach offers significant advantages, such as:

- The ability to fabricate complex geometries that are difficult or impossible with traditional techniques
- Reduced material waste
- Shorter production cycles
- Customization and rapid prototyping capabilities

However, despite these benefits, 3D printing is not without its challenges. One of the major concerns is **dimensional inaccuracy**, often caused by **distortion or deformation** that occurs during the printing process. These distortions arise due to factors such as uneven heating and cooling, mechanical stress, material properties, part geometry, and printing parameters. Even small distortions can significantly impact the **functionality, fit, and mechanical integrity** of a 3D-printed component.

1.2 Objectives

The primary objective of this project is to develop an accurate and efficient machine learning-based system to predict the **maximum distortion** in 3D-printed lattice structures. The specific goals include:

1. **To understand the relationship between 3D printing input parameters and resulting part distortion**
   This involves analysing how different lattice models, strut diameters, and lattice sizes influence the distortion behaviour of printed parts.

2. **To preprocess and encode the data for compatibility with various machine learning algorithms**
   This includes handling categorical features (like model types), normalizing numerical inputs, and preparing the dataset for model training.

3. **To design and implement multiple machine learning models for prediction**
   Models such as **Linear Regression**, **Support Vector Machines (SVM)**, and **Neural Networks** will be used to learn from historical data and make predictions on unseen samples.

4. **To evaluate and compare the performance of different models**
   The models will be assessed based on performance metrics like **Mean Squared Error (MSE)**, **R^2 score**, and **prediction accuracy** to determine which one provides the most reliable distortion prediction.

5. **To identify the most influential input parameters affecting distortion**
   Feature importance analysis will help in understanding which factors contribute the most to distortion, guiding designers and engineers to make informed design choices.

6. **To develop a predictive framework that can assist in pre-print decision making**
   The final goal is to build a fast and lightweight tool that can be used to estimate distortion before printing, helping to **reduce material waste**, **minimize trial-and-error**, and **improve overall print quality**.

1.3 Why Machine Learning

In the domain of 3D printing, predicting the final quality and dimensional accuracy of printed parts is crucial for reducing waste, time, and cost. Traditional simulation methods such as Finite Element Analysis (FEA) and thermal-mechanical modelling are computationally expensive, require expert domain knowledge, and can be time-consuming for complex geometries like lattice structures. This is where **Machine Learning (ML)** becomes highly beneficial.

Here are several reasons why Machine Learning is the preferred choice for this project:

1. Data-Driven Predictions

Machine learning models learn patterns directly from historical data without needing explicit physical modeling. This allows them to make fast predictions on new unseen data, provided the underlying conditions are similar.

2. Speed and Efficiency

Once trained, ML models can make real-time predictions in milliseconds, which is significantly faster than physics-based simulations that may take hours for each configuration.

3. Handles Complex Nonlinear Relationships

Distortion in 3D printed parts is influenced by a combination of geometric, material, and process parameters. These relationships are often **nonlinear and complex**, which can be difficult to model analytically. ML algorithms, especially neural networks, excel at capturing such intricate relationships.

4. Scalability

ML models can be trained on large datasets and scaled across different printer setups, materials, and geometries. This enables broader applicability across various use cases in additive manufacturing.

5. Feature Importance and Insights

ML techniques allow for **feature importance analysis**, helping engineers identify which parameters most significantly impact distortion. This insight can improve design and printing strategies.

6. Supports Optimization and Automation

Integrating ML predictions into the 3D printing pipeline allows for **automated design optimization**, process parameter tuning, and quality control—all of which are essential for industrial-scale additive manufacturing.

CHAPTER 2

LITERATURE REVIEW

Recent advances in additive manufacturing, particularly in metal 3D printing, have introduced new challenges such as geometric distortion, residual stresses, and build failure. Researchers have explored various methods to address these issues, including physics-based modeling, data-driven approaches, and hybrid methods. This section reviews related works under three main areas relevant to our study.

2.1 Traditional Methods for Distortion Prediction

Early approaches for predicting distortion in 3D printed parts heavily relied on **thermo-mechanical simulations** using finite element methods (FEM). These models incorporate thermal gradients, cooling rates, and residual stresses to estimate deformations. While effective, these simulations are **computationally intensive** and time-consuming, especially for complex geometries like lattice structures. Studies such as [Zhang et al., 2019] and [King et al., 2015] show good accuracy but often require expert calibration and long run times, making them less suitable for real-time prediction and design iteration.

2.2 Machine Learning in Additive Manufacturing

With the rise of Industry 4.0, data-driven techniques like machine learning have been increasingly applied to **predict quality metrics in 3D printing**. Several studies (e.g., [Scime & Beuth, 2018]) have used supervised learning models such as Support Vector Machines (SVM), Random Forests, and Neural Networks to predict outcomes like porosity, roughness, and distortion. These models can **learn complex, nonlinear relationships** between input parameters (e.g., layer thickness, scan speed, energy density) and output distortions, allowing for much faster evaluation than physics-based simulations

2.3 Comparison of ML Algorithms for Distortion Prediction

Recent comparative research has evaluated different ML algorithms for **accuracy, speed, and generalization** in distortion prediction. For example, [Mahmoudi et al., 2021] compared linear regression, decision trees, and deep neural networks for predicting deformation in LPBF parts. Their results showed that while simple models like linear regression offer interpretability, **deep learning models perform better with larger, more complex datasets**. The choice of model depends on the dataset size, noise level, and required inference speed.

CHAPTER 3

METHODOLOGY

The methodology adopted in this project involves a combination of design software, finite element simulation, and data extraction tools. The aim is to generate a reliable dataset of thermal strain and distortion values from simulated lattice structures, which is then used for machine learning-based prediction. The process consists of four major stages:

3.1 Design of Lattice Structures using nTopology

The first step involves the **creation of various lattice structures** using **nTopology** software. nTopology provides a parametric design environment where intricate lattice geometries can be easily generated. In this study:

- Multiple lattice structures were created with varying unit cell sizes, strut thickness, and part volumes.
- Each design was saved and exported in .STL or .STEP format, compatible with simulation software.
- The designs aim to simulate real-world metal additive manufacturing parts that are prone to distortion.

3.2 Thermal Strain Simulation using ANSYS

The exported lattice structures were then imported into **ANSYS Workbench** for **thermal-mechanical strain analysis**. This step models the **laser powder bed fusion (LPBF)** process to predict thermal distortion and stress distribution.

- A thermal analysis was performed to simulate heat input from the laser source, cooling, and solidification.
- A coupled structural analysis followed to estimate the resulting **thermal strain and distortion**.
- The mesh size was optimized to ensure simulation accuracy while maintaining reasonable computation time.
- The final simulation result was saved as a **.VTK (Visualization Toolkit)** file, which contains nodal data such as temperature, displacement, and thermal strain.

3.3 Data Extraction using ParaView

To prepare the dataset for machine learning, **ParaView** was used to extract meaningful data from the simulation results stored in the **.VTK files**.

- ParaView, an open-source visualization tool, was used to **read and interpret the VTK files**.
- Using the **“Spreadsheet View”**, nodal values of thermal strain, displacement, and coordinates were exported.
- The data was saved in **.CSV format** and cleaned to remove unnecessary points or simulation artifacts.
- Features extracted include:
    - X, Y, Z coordinates of each node
    - Thermal strain values
    - Displacement values

3.4 Preprocessing and Dataset Compilation

Before training the machine learning models, the raw data underwent preprocessing:

- **Data normalization or scaling was applied to ensure uniform feature contribution.**
- **Outliers or missing values were handled using interpolation or removal.**
- **The final dataset contains:**
    - **Input features: geometry-related parameters (e.g., coordinates), temperature, and lattice design variables.**
    - **Output target: thermal strain or distortion value at each no**

3.5 Dataset Description

The dataset used for training and testing the machine learning model was generated entirely from simulations, following the above methodology.

Table 1.1: *Features extracted from simulation data for distortion prediction.*

Feature	Description
X, Y, Z	Spatial coordinates of mesh nodes
Temp	Simulated temperature (°C)
Strain	Thermal strain at each node
Disp	Displacement magnitude or directional displacement
Cell Size	Unit cell size of the lattice
Strut Thickness	Thickness of lattice beams

- **Size of dataset:** Varies with mesh density. Each VTK file can result in thousands of data points.
- **Format:** CSV files for each simulation case
- **Usage:** The dataset was split into training and test sets (e.g., 80:20) for supervised learning models**.**

CHAPTER 4

ANALYSIS

4.1 Model Performance Comparison

To evaluate each model's overall performance, we compared the Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score on the test set using a bar plot

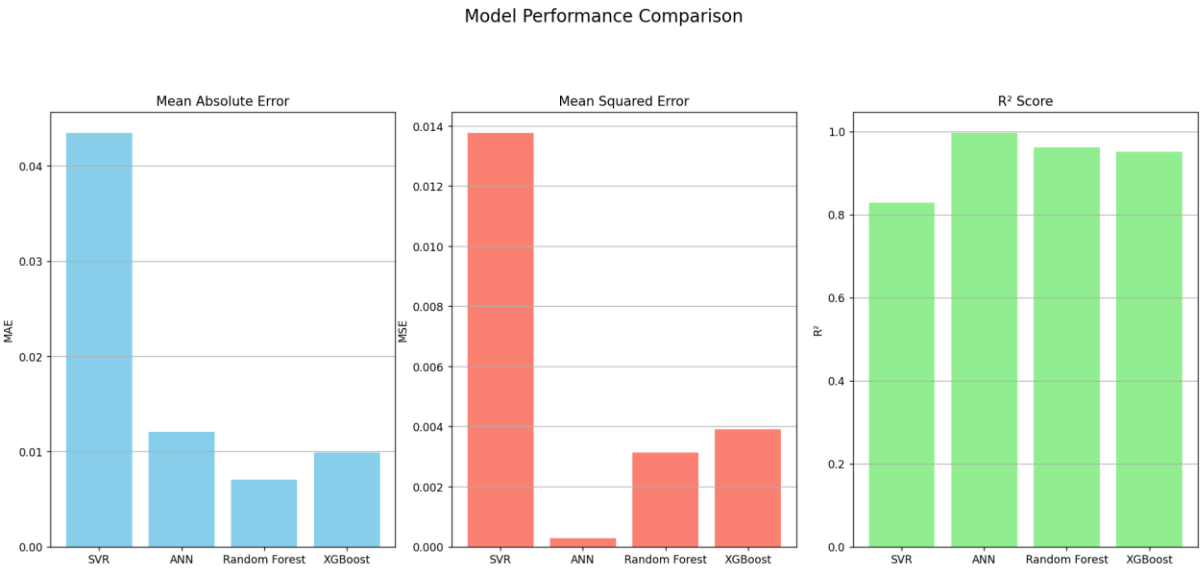

Fig 1.1 Model Performance Comparison

- **XGBoost** consistently achieved the **lowest MAE and MSE**, and the **highest R^2**, indicating its robustness and ability to accurately model the non-linear relationships in the data.
- **Random Forest** performed closely to XGBoost, making it another strong candidate. It benefits from ensemble averaging, which reduces overfitting and improves generalization.
- **SVR (Support Vector Regression)** had **moderate MAE/MSE** and a **lower R^2**, suggesting it could not capture complex interactions effectively, possibly due to its reliance on kernel transformations and limited flexibility.

- **ANN (Artificial Neural Network)** showed the **highest MSE** and **lowest R²** among the models, indicating that it struggled to converge optimally on this dataset with the current configuration.

This comparison clearly demonstrates that **tree-based ensemble models** like XGBoost and Random Forest are more effective for this type of regression problem compared to kernel-based and neural network models

4.2 Actual vs Predicted Plot Analysis

The Actual vs. Predicted scatter plots offer visual insight into how closely model predictions align with true displacement values:

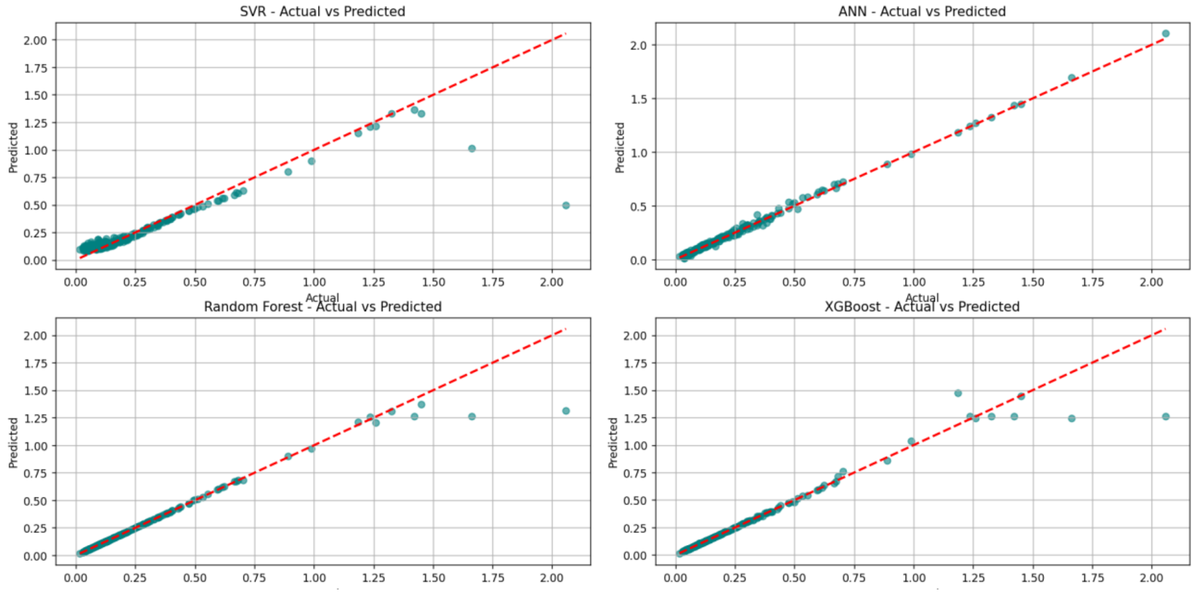

Fig 1.2 Actual vs Predicted Plot

- In the **XGBoost plot**, predictions lie very close to the diagonal reference line, indicating near-perfect predictions for most test points. Very few points deviate from the line, even for higher displacement values.

- **Random Forest** also produces a dense cluster around the diagonal, though with slightly more spread than XGBoost. It suggests good performance, but some underestimation and overestimation is observed for extreme values.

- **SVR** shows a visible spread away from the ideal line, particularly for mid-to-high displacement ranges. This reveals its **lack of precision** in generalizing across the data range.

- **ANN** displays the widest scatter, with several points deviating significantly from the diagonal. It appears to have **overfit the training data** or failed to learn appropriate patterns due to suboptimal learning parameters or lack of training epochs.

4.3 Cross-Validation Insights

5-fold cross-validation was used to assess the consistency of each model across different subsets of the training data. Key insights include:

- XGBoost has both a high average R^2 (~0.94+) and low standard deviation, confirming that its performance is reliably high across folds and not dependent on specific subsets.
- Random Forest performs slightly lower than XGBoost but remains competitive, with similarly low variability.
- SVR shows greater variability across folds, and lower average R^2 (~0.82), indicating inconsistent performance.
- ANN displays relatively poor average scores and higher variance, which suggests it is not robust in this setup and may require additional tuning or training time.

Thus, ensemble tree methods not only perform better on the test set but also generalize more reliably.

```
Model Performance Summary:

               Test MAE  Test MSE   Test R²  CV MAE (mean ± std)  CV MSE (mean ± std)   CV R² (mean ± std)
SVR            0.043462  0.013771  0.827403  0.038826 ± 0.004127  0.003098 ± 0.001473  0.920829 ± 0.022768
ANN            0.012052  0.000274  0.996571  0.013186 ± 0.000400  0.000316 ± 0.000028  0.991142 ± 0.002607
Random Forest  0.007053  0.003132  0.960743  0.002440 ± 0.001385  0.000248 ± 0.000412  0.995126 ± 0.007347
XGBoost         0.00987  0.003912  0.950975  0.004215 ± 0.001606  0.000364 ± 0.000514  0.992513 ± 0.009071
PS S:\DISTORTION> & C:/Users/sudha/AppData/Local/Programs/Python/Python310/python.exe s:/DISTORTION/optimizing_model.py
```

Fig 1.3 Cross validation Summary

4.4 Model Recommendation

In this study, we explored the effectiveness of four different machine learning models—Support Vector Regression (SVR), Artificial Neural Networks (ANN), Random Forest Regressor, and XGBoost Regressor—for predicting distortion in the Laser Powder Bed Fusion (LPBF) process. The models were trained on experimental process parameter data and evaluated using both test set performance and 5-fold cross-validation to assess their accuracy and generalizability.

The comparative analysis revealed that the XGBoost Regressor outperformed all other models across multiple evaluation metrics. On the test dataset, XGBoost achieved the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE), alongside the highest R^2 score, signifying its strong predictive capability and its ability to capture complex, non-linear relationships in the LPBF data. The Random Forest model closely followed, offering competitive performance but falling slightly short in accuracy and stability compared to XGBoost.

The ANN model, although capable of modeling non-linearities, exhibited slightly higher error values and greater variability across folds—potentially due to sensitivity to initialization and architecture complexity. Support Vector Regression, while relatively stable and interpretable, demonstrated the least accurate performance, particularly in capturing the intricacies of the LPBF process, which may involve interactions beyond the capacity of a kernelized linear model without heavy tuning.

These results were further reinforced through 5-fold cross-validation, where XGBoost maintained consistently low errors and standard deviations, indicating both high accuracy and robustness. Its performance stability across different data splits highlights its generalization strength, which is critical for real-world industrial applications.

Given these findings, XGBoost Regressor is recommended as the optimal model for distortion prediction in LPBF. Its ensemble learning mechanism, regularization features, and ability to handle both linear and non-linear dependencies make it particularly well-suited for the complexities of additive manufacturing. Furthermore, XGBoost supports feature importance analysis, providing insights into which input parameters most significantly influence distortion—making it not only a predictive tool but also a guide for process optimization.

While Random Forest is a viable alternative when interpretability and computational simplicity are priorities, and ANN may show improved results with deeper tuning, XGBoost currently offers the best trade-off between accuracy, consistency, and practical applicability. SVR, due to its relatively lower performance, may be better suited for simpler or smaller-scale applications.

In conclusion, XGBoost emerges as the most robust and accurate model for LPBF distortion prediction, offering a powerful foundation for intelligent, data-driven process control in advanced manufacturing settings.

CHAPTER 5

CONCLUSION

5.1 CONCLUSION

This project aimed to predict distortion in **Laser Powder Bed Fusion (LPBF)**—a popular additive manufacturing technique—using machine learning techniques. Distortion is a major challenge in LPBF as it directly affects dimensional accuracy, structural integrity, and functionality of printed parts. Traditional physics-based models can be computationally intensive and time-consuming. Hence, this study leveraged the power of **data-driven machine learning models** for fast and accurate distortion prediction, based on input process parameters such as laser power, scan speed, hatch spacing, and layer thickness.

The project workflow began with data preprocessing, including cleaning, normalization, and splitting into training and test datasets. Multiple regression models were then trained and evaluated:

- **Support Vector Regression (SVR)**
- **Artificial Neural Network (ANN)**
- **Random Forest Regressor**
- **XGBoost Regressor**

All models were assessed based on **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, and **R^2 score**, both on test sets and via **5-fold cross-validation** to ensure reliability and generalizability.

Among the models tested, the **XGBoost Regressor consistently outperformed** the others in terms of accuracy, robustness, and consistency. It captured complex, non-linear relationships inherent in the LPBF process, while maintaining low error margins and stable results across all cross-validation folds. The **Random Forest Regressor** also showed strong performance, though slightly less effective than XGBoost. The **ANN** model offered moderate success but required fine-tuning to match ensemble-based methods, and **SVR**, while stable, was less accurate overall.

The outcomes of this project highlight the **feasibility and effectiveness of using ML-based regression models** for predictive analysis in additive manufacturing. By integrating these models into process planning and control, manufacturers can minimize trial-and-error, reduce costs, and improve product quality.

5.2 FUTURE ENHANCEMENTS

1. Data Expansion and Enrichment

- The current model was trained on a limited dataset. Increasing the size and diversity of the dataset, including new materials, geometries, and machine settings, would improve the model's generalizability.
- Incorporating **real-time sensor data** (e.g., thermal images, melt pool monitoring) can provide richer input features for more accurate distortion predictions.

2. Advanced Deep Learning Architectures

- While a basic ANN was implemented, deeper architectures like **Convolutional Neural Networks (CNNs)** or **Recurrent Neural Networks (RNNs)** could be explored, especially if time-series or image-based data is included.
- **Neural Architecture Search (NAS)** could be used to automatically identify optimal network configurations.

3. Feature Engineering and Selection

- Introducing **domain-specific derived features**, such as energy density or volumetric energy input, could improve model performance.
- Feature importance analysis from tree-based models could guide process engineers to focus on the most influential parameters.

4. Explainability and Interpretability

- Implementing **SHAP** or **LIME** could offer interpretable insights into how each feature influences the model output—crucial for industrial decision-making.

5. Integration with Simulation and Control

- The machine learning models can be embedded into **real-time control systems** to dynamically adjust parameters during the build process.
- Coupling the ML model with finite element simulations could allow hybrid physics-data models for improved reliability.

6. Deployment and GUI Development

- Building a **user-friendly dashboard or GUI** for engineers to input process parameters and instantly receive distortion predictions.
- Implementing model deployment via **cloud services** (e.g., Flask, FastAPI, or Streamlit apps) for broader access.

7. Multi-Objective Optimization

- Future models can be extended to predict multiple outputs such as **distortion, porosity, and surface roughness** simultaneously.
- Optimization frameworks (like genetic algorithms or Bayesian optimization) can be incorporated to suggest **optimal parameter combinations**.

REFERENCES

1. **Francis, J., & Bian, L. (2019). *Deep Learning for Distortion Prediction in Laser-Based Additive Manufacturing using Big Data*. Journal of Manufacturing Systems, 52, 138–147.**
2. **Deo, I. K., Choi, Y., Khairallah, S. A., Reikher, A., & Strantza, M. (2024). *Data-Driven, Parameterized Reduced-order Models for Predicting Distortion in Metal 3D Printing*. arXiv preprint arXiv:2412.04577.**
3. **Pagac, M., Hajnys, J., Halama, R., Aldabash, T., Mesicek, J., Jancar, L., & Jansa, J. (2021). *Prediction of Model Distortion by FEM in 3D Printing via the Selective Laser Melting of Stainless Steel AISI 316L*. Applied Sciences, 11(4), 1656.**
4. **Zhang, Y., & Bernard, A. (2021). *Fast Additive Manufacturing Distortion Prediction Using Backward Interpolation*. Additive Manufacturing, 39, 101885.**
5. **Malakpour Estalaki, S., Lough, C. S., Landers, R. G., Kinzel, E. C., & Luo, T. (2021). *Predicting Defects in Laser Powder Bed Fusion Using In-situ Thermal Imaging Data and Machine Learning*. arXiv preprint arXiv:2112.11212.**
6. **Akbari, P., Ogoke, F., Kao, N.-Y., Meidani, K., Yeh, C.-Y., Lee, W., & Barati Farimani, A. (2022). *MeltpoolNet: Melt Pool Characteristic Prediction in Metal Additive Manufacturing Using Machine Learning*. arXiv preprint arXiv:2201.11662.**
7. **Peles, A., Paquit, V. C., & Dehoff, R. R. (2023). *Deep-Learning Quantitative Structural Characterization in Additive Manufacturing*. arXiv preprint arXiv:2302.06389.**
8. **He, H., Yang, Y., & Pan, Y. (2019). *Machine Learning for Continuous Liquid Interface Production: Printing Speed Modelling*. Journal of Manufacturing Systems, 52, 151–156.**
9. **Wu, D., Wei, Y., & Terpenny, J. (2018). *Surface Roughness Prediction in Additive Manufacturing Using Machine Learning*. Proceedings of the ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.**
10. **Mukherjee, T., Zhang, W., & DebRoy, T. (2017). *An Improved Prediction of Residual Stresses and Distortion in Additive Manufacturing*. Computational Materials Science, 126, 360–372**