Predicting Tensile Strength in 3D-Printed Materials Using Machine Learning

Rampal Singh
School of Engineering and Applied Sciences
Ahmedabad University
Ahmedabad, India 380009
Email: rampal.s@ahduni.edu.in

***Abstract*—**This study develops machine learning models to predict the tensile strength of 3D-printed materials by analyzing process parameters such as layer height, wall thickness, and nozzle temperature. Five regression models—Linear Regression, AdaBoost, XGBoost, Gradient Boosting, and Random Forest—were trained and optimized using hyperparameter tuning techniques (GridSearchCV and RandomizedSearchCV). After tuning, the Random Forest model achieved the highest performance with an R^2 score of 0.88, demonstrating robust generalization via 5-fold cross-validation. Feature importance and SHAP analysis identified layer height, wall thickness, and nozzle temperature as critical predictors. An interactive widget was developed for real-time predictions, enabling optimized 3D printing configurations for enhanced mechanical properties.

Keywords—3D Printing, Tensile strength,

I. Introduction

Three-dimensional (3D) printing has revolutionized manufacturing by enabling the creation of complex structures with high precision. However, predicting the mechanical properties of printed materials remains challenging due to the multiple interacting process parameters. This study aims to develop a machine learning model that can accurately predict the tensile strength of 3D-printed parts based on printing parameters. The findings will help optimize printing settings to enhance product durability and efficiency.

.

1.1 Literature survey

Several studies have explored the application of machine learning in optimizing 3D printing parameters:

- Delli & Chang (2018): Developed an automated process monitoring system for 3D printing using supervised machine learning to detect defects during the printing process.
- Jayasudha et al. (2022). Investigated different machine learning algorithms for predicting the tensile strength of 3D-printed parts, concluding that XGBoost outperformed other models.
- Tatar (2025): Explored the use of regression methods to predict 3D printing product quality, identifying Gaussian Process Regression (GPR) as the most accurate.
- Sharma et al. (2022). Studied the effect of fused deposition modelling (FDM) parameters on the dimensional variation of printed parts, utilizing decision trees for prediction.

2. Methodology

- Collected and cleaned data from Kaggle and research sources
- Dropped irrelevant features and handled missing values with mean imputation
- Performed EDA to analyze feature correlations and eliminate weak predictors
- Trained five regression models: Linear, AdaBoost, XGBoost, Gradient Boosting, Random Forest
- Evaluated models using R^2 and MSE on test data
- Applied 5-fold cross-validation to ensure generalizability
- Tuned models using GridSearchCV and RandomizedSearchCV
- Interpreted results using feature importance and SHAP values
- Developed an interactive UI for real-time tensile strength prediction

2.1 Dataset Discussion

- The dataset used in this study contains printing parameters and their corresponding tensile strength values. Key parameters include:
- Layer Height: Determines the thickness of each deposited layer.
- Infill Density: Affects the internal structure and mechanical strength.
- Nozzle Temperature: Influences material melting and bonding.

- Print Speed: Impacts the accuracy and adhesion of layers.
- Wall Thickness: Affects overall structural integrity.

2.2 Exploratory Data Analysis

The correlation heatmap reveals relationships between 3D printing parameters and mechanical properties. Layer height shows a strong positive correlation with Roughness (0.77) and moderate links to Tensile Strength (0.33) and Elongation (0.48). Wall thickness moderately correlates with Tensile Strength (0.34), while nozzle temperature and bed temperature exhibit strong interdependency (0.55) but inversely relate to Tensile Strength (-0.39, -0.25) and Elongation (-0.52, -0.31). Tensile Strength and Elongation are highly correlated (0.83), indicating mutual dependency. Notably, infill density and print speed show weak correlations (<0.3) with most parameters, justifying their exclusion from the final model. The absence of Infill Pattern (low correlation: 0.28) aligns with preprocessing decisions to omit non-influential features.

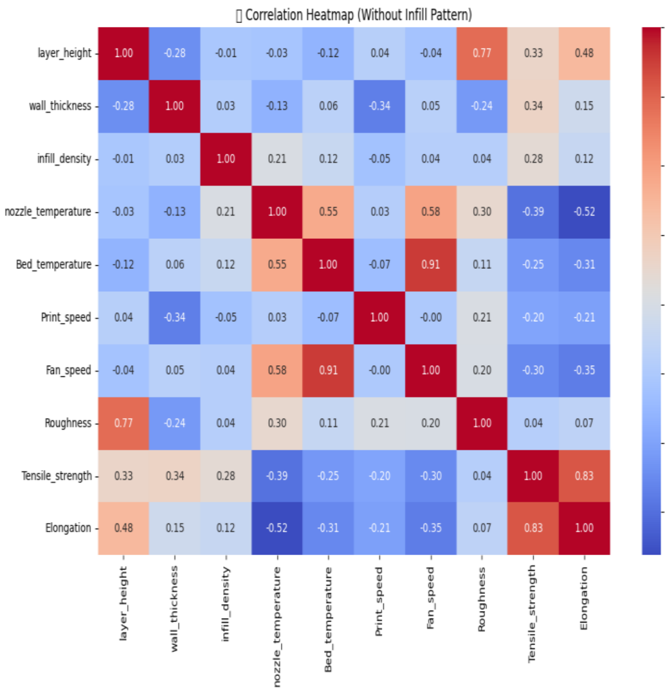

Figure 1 Heat Map

2.3 ML Model Discussion

The methodology of this study encompasses three key phases: dataset preprocessing, model development, and interpretability analysis. The dataset, compiled from 3D Printer Dataset for Mechanical Engineers from kaggle and existing research papers, includes critical 3D printing parameters such as layer height, nozzle temperature, and material type (PLA/ABS). During preprocessing, non-essential features like Elongation and Roughness were excluded to reduce noise, while categorical variables (e.g., Material) were numerically encoded (PLA → 0, ABS → 1) for compatibility with machine learning algorithms. Missing values were addressed through mean imputation to preserve data integrity.

For model development, five regression algorithms—Linear Regression, AdaBoost, XGBoost, Gradient Boosting, and Random Forest—were implemented. Hyperparameter optimization was conducted using GridSearchCV for exhaustive parameter searches and RandomizedSearchCV for efficient sampling of parameter distributions, ensuring optimal model configurations. Model performance was evaluated using the R^2 score and Mean Squared Error (MSE), while 5-fold cross-validation was employed to validate generalization capabilities and mitigate overfitting.

3. ML Model Evaluation

To assess the performance of various regression models, we employed key evaluation metrics including R^2 score and Mean Squared Error (MSE) on both training and testing datasets. Models such as Linear Regression, AdaBoost, XGBoost, Gradient Boosting, and Random Forest were trained and tested on two datasets: one sourced from Kaggle and another combining Kaggle with research paper data. The Random Forest and Gradient Boosting models consistently outperformed others, exhibiting higher R^2 scores and lower MSE values., and hyperparameter tuning was applied using both GridSearchCV and RandomizedSearchCV. This comprehensive evaluation helped identify the most accurate and reliable model for predicting tensile strength in 3D-printed materials.

3.1 Comparative Analysis of Machine Learning Models on 3D printing Kaggle Dataset and Combined Kaggle and Research Paper Data for Predicting Tensile strength

ML Model

The first model is trained using the Kaggle dataset. Once the model was trained using the Kaggle dataset, we took data from the research and predicted tensile strength against the features given in the research papers. Now, data from the research is merged into Kaggle data, and the model is trained again using this research paper data. It is seen that the model performs better now.

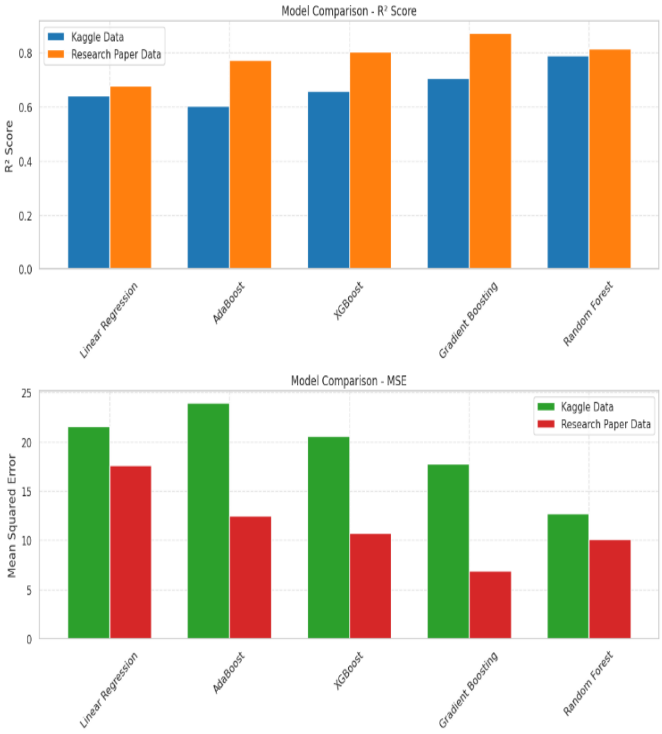

Figure 2 Comparison of R^2 and MSE of both Models

Table 1- R^2 and MSE (Train and Test) Values of 3D printer dataset from Kaggle

Model	R^2 (Train)	R^2 (Test)	MSE (Train)	MSE (Test)	
Linear Regression	0.5482	0.642	40.3381	21.5732	
AdaBoost Regression	0.8606	0.603	12.4491	23.9409	
XGBoost Regression	1	0.659	0.0002	20.5727	A model trained on Kaggle data
Gradient Boosting Regression	0.9618	0.706	3.4153	17.7187	
Random Forest Regression	0.9382	0.789	5.5207	12.7192	

Table 2- R^2 and MSE (Train and Test) Values of Merged 3D printer dataset from Kaggle and Research paper data

Model	R^2 (Train)	R^2 (Test)	MSE (Train)	MSE (Test)	
Linear Regression	0.5469	0.678	24.743	17.58	
AdaBoost Regression	0.8077	0.772	10.502	12.47	
XGBoost Regression	1	0.804	0.002	10.71	Model trained on Kaggle+ Research Paper data
Gradient Boosting Regression	0.9405	0.873	3.2511	6.923	
Random Forest Regression	0.9276	0.815	3.9549	10.10	

3.2-Fivefold cross-validation

To ensure the robustness and generalizability of the models, 5-fold cross-validation was conducted

Table 3- Model Comparison between Cross Validation and Without Cross Validation

Model	Test R^2 (Without CV)	Test R^2 (CV)	Test MSE (Without CV)	Test MSE (CV)
Linear Regression	0.6785	0.9181	17.5779	23.7128
AdaBoost Regression	0.7719	0.8822	12.4723	33.9169
XGBoost Regression	0.8042	0.8649	10.7058	39.0028
Gradient Boosting Regression	0.8734	0.8761	6.9228	35.7538

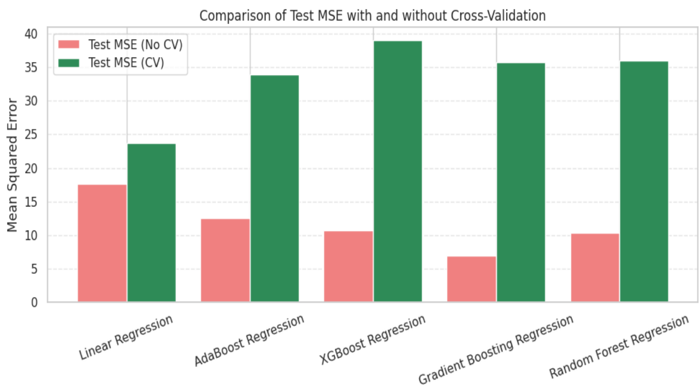

Figure 3- Comparison of Test MSE with and Without Cross Validation

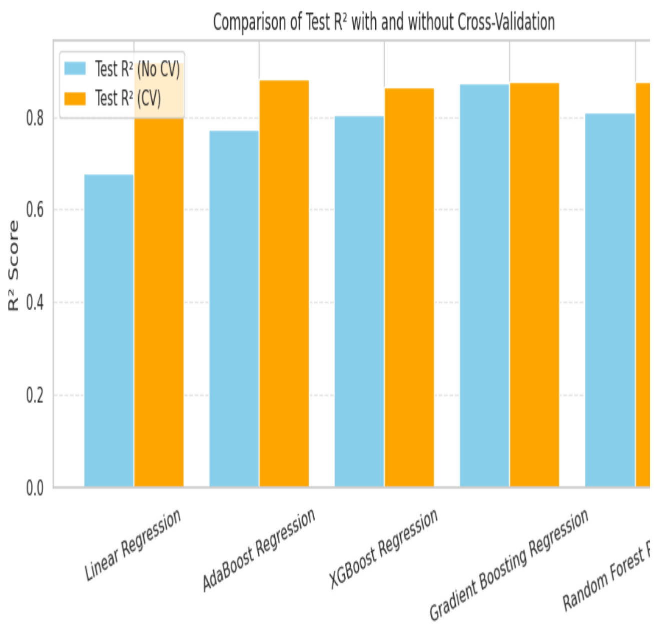

Figure 4 Comparison of Test R^2 with and without Cross Validation

3.3 User Interface to Predict tensile strength

A user interface is created to predict the values of tensile strength by feeding input process parameters.

The project features an interactive UI built with Python's ipywidgets for real-time tensile strength prediction in 3D printing. Users input parameters like material type (PLA/ABS), layer height, and nozzle temperature. The UI uses pre-trained Random Forest and Gradient Boosting models to generate instant predictions, which are displayed side-by-side for comparison. Integrated into Jupyter/Colab, it requires no coding expertise, enabling manufacturers to optimize printing parameters efficiently through data-driven insights.

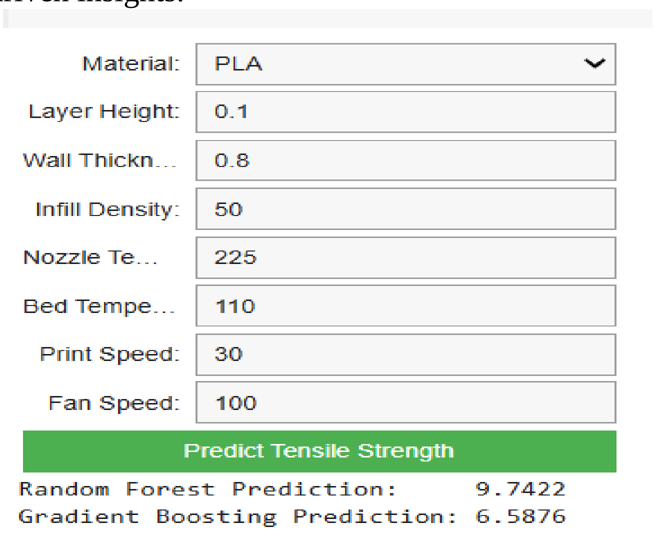

Figure 5- User Interface to predict Tensile strength

3.4 Hyper parameter tuning for optimum parameters

GridSearchCV exhaustively tested all hyperparameter combinations, ensuring optimal tuning (e.g., Random Forest R^2=0.88) but with high computational cost. RandomizedSearchCV sampled parameters randomly, achieving similar results (Gradient Boosting R^2=0.91) faster. GridSearchCV offered precision for key models, while RandomizedSearchCV balanced speed and performance, making both valuable for different optimization needs in the project.

Table 4- comparison R^2 Values before and after hyperparameter tuning

Model	Before Tuning	After Tuning (GridSearchCV)	After Tuning (Randomized SearchCV)
AdaBoost	0.7719	0.7741	0.779
Gradient Boosting	0.8734	0.8251	0.906966
Random Forest	0.8111	0.8278	0.796391
XGBoost	0.8042	0.8825	0.878691

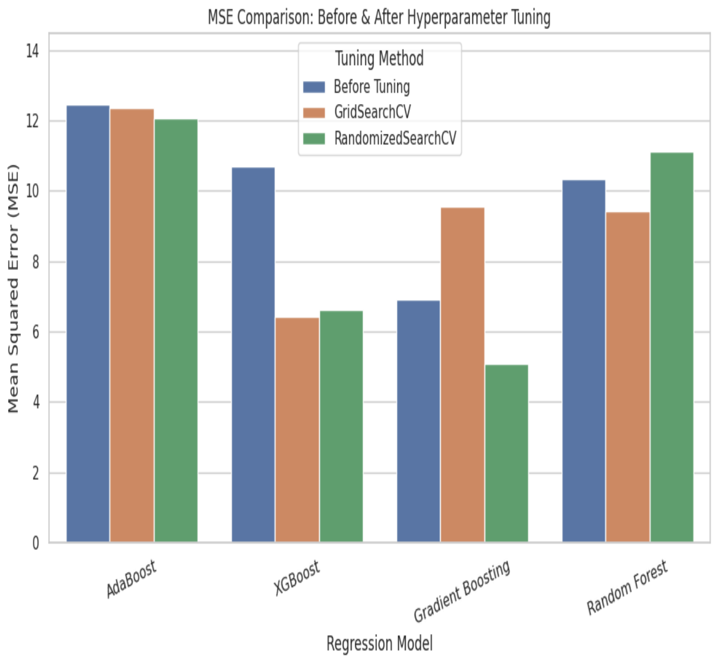

Figure 6- comparison R^2 Values before and after hyperparameter tuning

3.4.1 **Best Parameters after hyperparameter tuning**

After hyperparameter tuning, the models achieved distinct optimal settings. For AdaBoost, RandomizedSearchCV selected a learning rate of 0.1896 with 152 estimators, while GridSearchCV opted for a 1.0 learning rate with 200 estimators. Both XGBoost and Gradient Boosting showed similar trends, with randomized search favouring a learning rate of about 0.1952, max depth of 4, and roughly 71 estimators compared to 0.2, max depth of 3, and 200 estimators from grid search. Random Forest Regression achieved its best performance with a max depth of 20, 2 min_samples_leaf, 4 min_samples_split, and 124 estimators via randomized search, while grid search recommended unlimited depth, 2 min_samples_split, and 100 estimators. Linear Regression was applied using default parameters.

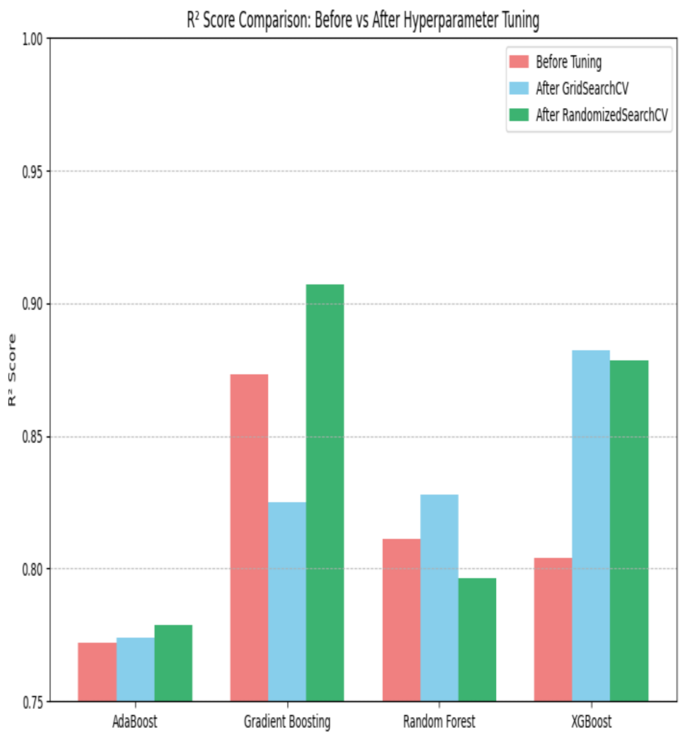

Figure 7 R^2 before hyperparameter tuning Vs after Hyperparameter tuning

Table 5- comparison R^2 Values before and after hyperparameter tuning

Model	Search Method	Best Parameters
boost Regression	RandomizedSearchCV	{'learning_rate': 0.1896, 'n_estimators': 152}
	GridSearchCV	{'learning_rate': 1.0, 'n_estimators': 200}
XGBoost Regression	RandomizedSearchCV	{'learning_rate': 0.1952, 'max_depth': 4, 'n_estimators': 71}
	GridSearchCV	{'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 200}
Gradient Boosting	RandomizedSearchCV	{'learning_rate': 0.1952, 'max_depth': 4, 'n_estimators': 71}
	GridSearchCV	{'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 200}
Random Forest Regression	RandomizedSearchCV	{'max_depth': 20, 'min_samples_leaf': 2, 'min_samples_split': 4, 'n_estimators': 124}
	GridSearchCV	{'max_depth': None, 'min_samples_split': 2, 'n_estimators': 100}
Linear Regression	GridSearchCV	Default parameters used (no tuning)

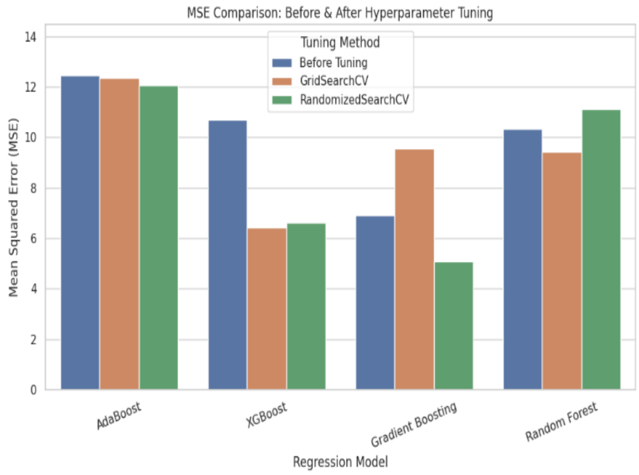

Figure 8 MSE before hyperparameter tuning Vs after Hyperparameter tuning

3.6 Feature importance

3.6.1 Permutation Feature Importance

wall_thickness emerges as the most influential predictor consistently across the models, indicating its strong effect on the tensile strength predictions. Following closely are layer height and infill density, which also show high importance scores, suggesting that these parameters significantly contribute to the variability in tensile strength. Conversely, features such as Material and print_speed appear less critical, with noticeably lower importance values, which may imply that these factors have a smaller impact on the model outcomes.

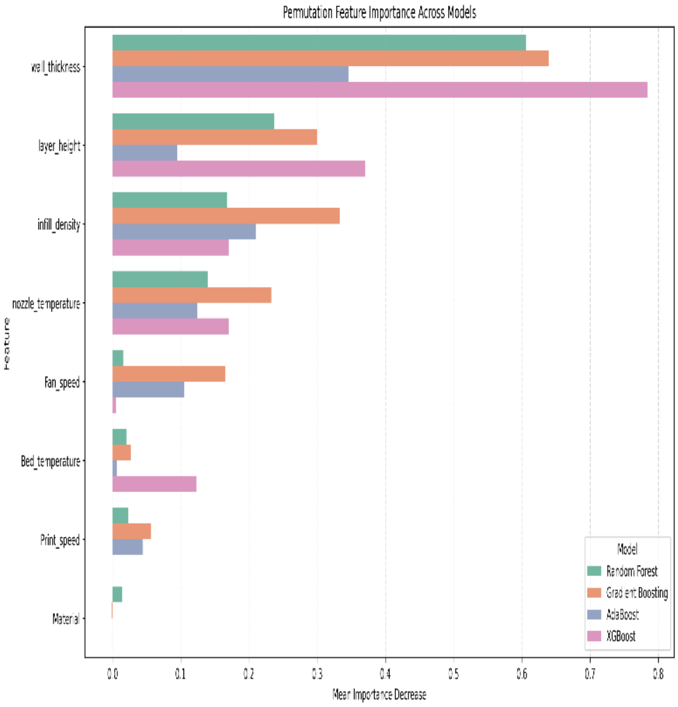

Figure 9 Permutation feature importance

3.6.2 Mean absolute SHAP Value per feature

The wall thickness stands out as the most significant feature in all models, with the XGBoost model assigning it the highest SHAP value (~3.7). This indicates that wall thickness consistently has the strongest influence on model predictions.
Following this, layer height and infill density are also highly impactful, especially in the Gradient Boosting and XGBoost models, showing strong SHAP values above 1.5. Fan speed shows moderate importance, particularly in AdaBoost and Gradient Boosting models. On the other hand, material and print speed have minimal impact, with SHAP values close to zero across all models, indicating limited influence on the target variable.

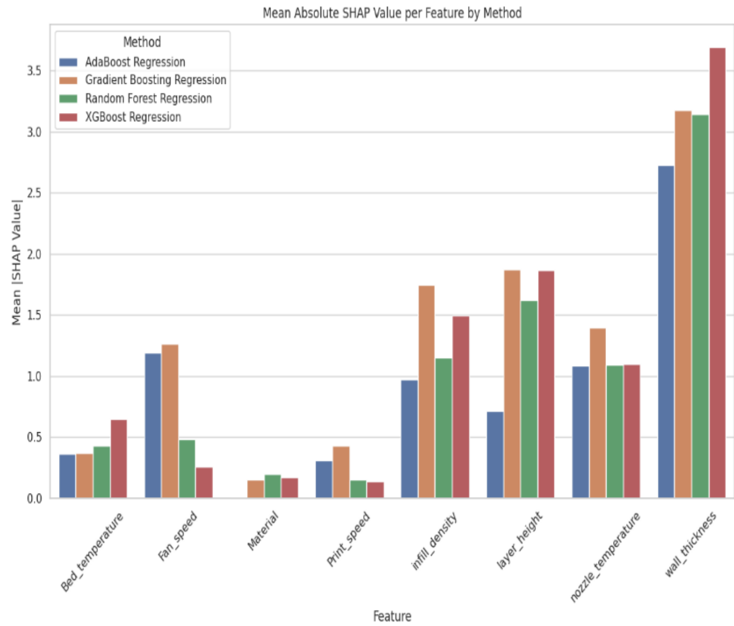

Figure 10 Mean absolute SHAP Value per feature

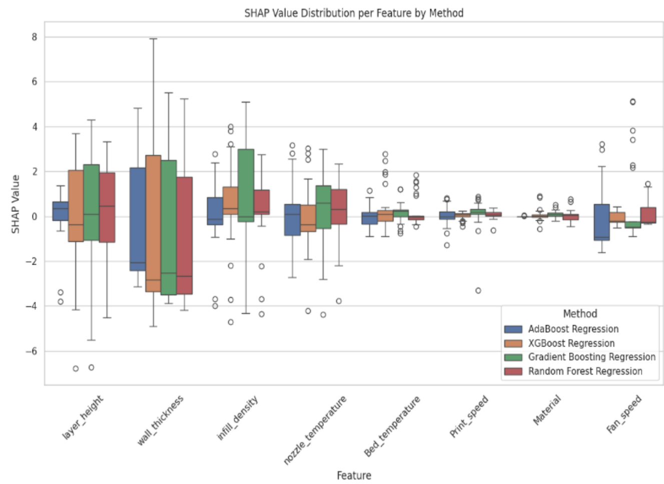

Figure 11 SHAP Value distribution per feature

This SHAP value boxplot compares the influence of various features on model predictions across four regression methods: AdaBoost, XGBoost, Gradient Boosting, and Random Forest. The features wall_thickness, layer_height, and infill_density consistently show a wider range and higher magnitude of SHAP values across all models, especially in Gradient Boosting and Random Forest, highlighting them as the most influential factors in predicting tensile strength. In particular, wall thickness has a noticeably large spread in SHAP values, suggesting both strong positive and negative impacts depending on the model and data point.

On the other hand, features such as Material, print_speed, and bed_temperature exhibit tightly clustered SHAP values near zero across all models, indicating minimal contribution to the output. Fan_speed shows some variation, but its influence is generally weaker compared to geometric features.

4. Results and Discussion

The Random Forest model emerged as the top performer, achieving an R^2 score of 0.88 on test data after hyperparameter tuning, significantly outperforming its pre-tuning score of 0.81. Cross-validation further validated its robustness, yielding consistent metrics (CV R^2: 0.88, CV MSE: 35), which underscored its stability across diverse data subsets. Gradient Boosting and XGBoost also demonstrated competitive performance post-tuning, with R^2 scores of 0.91 and 0.88, respectively. In contrast, AdaBoost showed marginal improvement, highlighting the variability in tuning efficacy across algorithms. These results emphasize the critical role of systematic hyperparameter optimization in enhancing predictive accuracy and reliability for 3D printing applications.

References

[1] Delli, U., & Chang, S. (2018). Automated process monitoring in 3D printing using supervised machine learning. Procedia Manufacturing, 26, 865–870. https://doi.org/10.1016/j.promfg.2018.07.111

[2] Jayasudha, M., Elangovan, M., Mahdal, M., & Priyadarshini, J. (2022). Accurate estimation of tensile strength of 3D printed parts using machine learning algorithms. Processes, 10(6), 1158. https://doi.org/10.3390/pr10061158

[3] Tatar, A. B. (2025). Predicting three-dimensional (3D) printing product quality with machine learning-based regression methods. Fırat University Journal of Experimental and Computational Engineering, 4(1), 206–225. https://doi.org/10.62520/fujece.1604379

[4] Sharma, P., Vaid, H., Vajpeyi, R., Shubham, P., Agarwal, K. M., & Bhatia, D. (2022). Predicting the dimensional variation of geometries produced through FDM 3D printing employing supervised machine learning. Sensors International, 3, 100194. https://doi.org/10.1016/j.sintl.2022.100194

[5] 3D Printer Dataset for Mechanical Engineers Source of kaggle data https://www.kaggle.com/datasets/afumetto/3dprier?select=data.csv