3D Bioprinting Printability & Error Prediction

Objective:

To build a predictive system that understands which parameters (like material composition, speed, and extrusion rate) influence the *Diameter Error* of printed samples, improving print quality and process optimization.

1. Feature Selection — Identifying the Key Input Variables

After cleaning and analyzing the dataset, we carefully selected the most relevant variables (features) that influence print quality.

Selected Input Features:

Feature	Description	Type
E/S	Extrusion-to-speed ratio (balance between material flow and motion)	Numeric
Extrusion Rate	How fast the material is extruded	Numeric
Speed	Printing speed	Numeric
Alginate_%, **HAMA_%**, **GelMA_%**	Material composition percentages	Numeric
Needle Diameter_27G	Encoded representation of needle size	Categorical (One-hot)
Crosslinker_CaCl2 + Ru/SPS	Encoded representation of crosslinker type	Categorical (One-hot)

Why these were chosen?
These parameters directly affect layer formation, print shape retention, and final geometry.
Noise columns like *Image name*, *ID*, and *printability* were excluded to prevent model bias or data leakage.

2. Data Scoping — Handling the *Printability* Flag

The **printability** flag represents whether a print attempt was successful (1) or not (0).

To ensure data consistency:

> For **regression tasks** (predicting *Diameter Error*), we **only used rows where** printability = 1, i.e., successful prints.
>
> For **classification tasks** (future work), the entire dataset including unsuccessful prints is retained.
>
> This ensures that the model learns *only from successful printing cases*, avoiding distortions from failed or incomplete prints.

3. Model Enhancement — Adding & Optimizing XGBoost

We compared three models to predict *Diameter Error*

Evaluation table:

Model	**Training R^2**	**Testing R^2**	**MAE (Training)**	**MAE (Testing)**	**RMSE (Training)**	**RMSE (Testing)**
Linear Regression	0.591	0.650	0.256	0.232	0.354	0.324
Random Forest Regressor	0.931	0.863	0.103	0.125	0.146	0.202
XGBoost Regressor	0.932	0.857	0.101	0.131	0.145	0.207
Tuned XGBoost (with GridSearchCV)	0.924	0.856	0.112	0.118	0.152	0.208

1. Linear Regression acted as the baseline, performing moderately with limited ability to capture complex, non-linear relationships.
2. Random Forest significantly improved accuracy, achieving over 86% R^2 on test data with much lower error rates.
3. XGBoost Regressor performed almost identically to Random Forest but with more computational efficiency.
4. After fine-tuning through GridSearchCV, the Tuned XGBoost model

achieved the best overall balance between accuracy and generalization.
5. The minimal gap between training and testing R^2 values indicates excellent model stability without overfitting.

In simple terms — the optimized XGBoost model can accurately predict Diameter Errors (~0.12 mm deviation), making it the most reliable choice for real-world 3D bioprinting applications.In simple terms, this means the model can reliably predict Diameter Errors with an average deviation of only about 0.12 mm, making it ideal for guiding print parameter optimization.

5.Clarity of Visualizations — Heatmap Improvements

We refined the **correlation heatmap** to improve interpretability:

Key Enhancements:

Used **"coolwarm" color palette** for better contrast between positive and negative correlations.

Numeric-only filtering to remove irrelevant columns.

Reordered columns to place the target (*Diameter Error*) last for easy visual focus.

Saved high-resolution image (correlation_heatmap.png) for reporting and presentations.