Artificial Intelligence and Machine Learning in Printed Electronics Design

A Design Project Report

Presented to the School of Electrical and Computer Engineering of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering, Electrical and Computer Engineering

Submitted by

Anzhou Li (al2627), Shizhe Shen (ss4335), Yusheng Chen (yc2832)

MEng Field Advisor: Peter Doerschuk

MEng Outside Advisor: Ben Davaji

Degree Date: May 2025

Abstract

Master of Engineering Program

School of Electrical and Computer Engineering

Cornell University

Design Project Report

Project Title: Artificial Intelligence and Machine Learning in Printed Electronics

Author: Anzhou Li (al2627), Shizhe Shen (ss4335), Yusheng Chen (yc2832)

Abstract

This project aims to develop data-driven models to map the relationship between geometric structure and electrical performance (resistance) in printed electronic devices, and to compare the predictive accuracy of models using target versus measured dimensions. Two datasets were used: R1, which includes only geometric parameters, and R2, which incorporates additional process variables such as temperature and Dynamic Top Oven (DTO) speed.

For both datasets, models were trained using target and measured inputs. In R1, a linear regression model using width and gap achieved the best performance, with minimal differences in error distribution between target and measured inputs, supporting model stability under low-complexity conditions. In R2, a second-order polynomial regression model with LASSO regularization was adopted to handle the expanded feature space. This model better captured the nonlinear interactions between process parameters and conductance, showing improved robustness, especially with measured inputs.

Overall, measured parameters demonstrated stronger resilience to printing variability. Linear models proved effective for simple cases, while polynomial regression offered better expressiveness for complex, multi-variable scenarios. The results provide a foundation for predictive modeling and process control in flexible printed electronics.

Anzhou Li's contribution

Anzhou Li's contribution includes project coordination, model development, and experimental analysis. He communicated with the project advisor to clarify the project direction and determine appropriate modeling approaches. He explored multiple regression models and evaluated their suitability for the dataset. Additionally, he contributed to the generation of 3D plots, slice plots, and umbrella plots, and contributed to poster preparation and interpretation of the final results and writing the final report.

Shizhe Shen's contribution

Shizhe worked on model selection, data processing, and validation. He used Weka to identify the best model for the R1 dataset and cleaned the R2 data in coordination with Haiyang. Using Scikit-learn, he trained and compared models—including linear regression, multilayer perceptron, multinomial regression, and random forest—then applied the best one to R1 for verification. He also supported project logistics through weekly reports, presentations, and final report writing.

Yusheng Chen's contribution

In this project, Yusheng mainly contributed to model development and testing, including machine learning model validation using Weka, as well as coding for model training with the Scikit-learn package. He also contributed to the creation of figures, the generation of 3D plots, and post-experimental analysis. In addition to his technical contributions, Yusheng was actively involved in project logistics, including preparing weekly progress reports and presentations for the project advisor, as well as writing portions of the final deliverables.

Executive Summary

This project addresses a key challenge in the field of printed electronics: the variability between designed and actual device performance due to process-induced deviations. These deviations—originating from factors such as inaccurate line widths, gap spacing, temperature fluctuations, and oven speed—can significantly affect electrical resistance and product yield. To tackle this, we developed and evaluated machine learning models capable of predicting electrical resistance based on both design (target) and actual (measured) geometric and process parameters.

The project uses two data sets: R1 contains geometric input based on layout only, and R2 introduces process parameters such as temperature and DTO. For each set of data, we have built prediction models based on target dimensions and measured dimensions respectively. The analysis results show that although the average prediction accuracy of the two types of inputs is similar, the model based on measured dimensions is more concentrated in error distribution, with fewer extreme errors, showing stronger robustness and anti-fluctuation ability.

This discovery has direct guiding significance for the quality control of flexible hybrid electronics (FHE) manufacturing. In the initial stage of modeling, reasonable performance can be obtained by using design parameters, but when entering the production process, the stability and reliability of the prediction model can be significantly improved by fusing real-time measurement data.

We recommend incorporating real-time measurement data into the prediction pipeline of printed electronics fabrication systems. This data-driven approach can reduce resistance variability, minimize material waste, and improve yield, all without requiring changes to hardware. The lightweight nature of our modeling framework also makes it practical for integration into existing low-cost FHE manufacturing workflows.

By improving prediction accuracy and robustness, this project contributes to smarter, more efficient manufacturing processes in the growing field of printed electronics.

Introduction

The research background of this project stems from the rapid development demand of flexible and hybrid electronics (FHE) technology in the printing electronics industry. With the emerging applications such as smart wearable devices, flexible sensors, medical patches and Internet of Things devices, FHE is gradually becoming an important supplement or even an alternative to traditional rigid electronic manufacturing with its manufacturing advantages of high customization, low cost and rapid iteration. Driven by this trend, the industry has put forward higher requirements for efficient, accurate and stable FHE printing solutions.

However, the current FHE printing process still faces many technical bottlenecks, which restrict its large-scale promotion and quality consistency control. The instability of printing quality is one of the most important problems, which is manifested in the large fluctuation of device line width and spacing, resulting in the deviation of resistance, capacitance and other performance from the design value. Nozzles are frequently blocked, which seriously affects production continuity and equipment life. In addition, the lack of predictability of the final electrical performance makes it difficult for engineers to judge the actual performance of finished products through the preliminary design parameters, resulting in high trial-and-error costs and serious waste of materials.

Based on the above problems, this project proposes to build an optimization modeling framework driven by artificial intelligence, aiming at fundamentally improving the forecasting ability, stability and automation level of the FHE printing process. Based on the actual measurement data and regression model technology, the framework will establish the mapping relationship between design parameters, printing process parameters and final electrical properties (such as resistance), to realize high-precision prediction and error control of finished product performance.

Through this research, we hope to provide a sustainable and reproducible intelligent data modeling paradigm for the future FHE manufacturing system and promote printing electronics from the laboratory to the standardized and large-scale production stage.

Design Problem and Requirements

I. Design Objectives

The main design goals of this project are: firstly, to build a high-precision and low-cost data-driven prediction model, which can accurately predict the geometric structure (such as line width and spacing) and electrical properties (such as resistance) of printed electronic devices. Secondly, reduce the cost and resource consumption of experiments by reducing unnecessary physical printing verification. Thirdly, it provides an intelligent decision-making tool for the FHE manufacturing process, which can be used as a key support system for parameter optimization, yield improvement and quality control.

II. System Requirements

In order to achieve the above objectives, the designed system should meet the following functional and platform compatibility requirements. First, the system should be able to receive a variety of inputs, including design layout parameters (such as line width and spacing), process parameters (such as temperature, baking speed DTO, nozzle pressure, etc.) and optional environmental factors (such as humidity). The second digital twin simulation function can predict the geometric and resistance characteristics of printed results without actual printing operation, and support the "virtual printing experiment" based on historical data for parameter adjustment, early warning and optimization. Third, it is compatible with the printing equipment platform. The input and output format of the model should be compatible with the data interface of the NanoDimension DragonFly IV 3D printing platform.

III. Design Constraints

Although the project objectives are clear, it still faces the following practical limitations in the process of design and implementation:

i. First, the training data is limited. At present, the sample size of the data set is small and the feature dimension is limited, which may limit the complexity and generalization ability of the model and needs to be compensated by means of data enhancement and regularization.
ii. Second, experimental resources are limited. The limited use time and test quantity of printing equipment in the laboratory limit the iterative verification ability of the model on real equipment, and it needs more off-line analysis and simulation verification.
iii. Third, at the same time, the model needs to take into account the computational efficiency and deployment feasibility, and avoid over-reliance on the cloud or

high-computing servers.

iv. Fourth, time and resource constraints. The project cycle is one year, and data processing, modeling, error analysis, system testing and document writing need to be completed in a limited time, so the design needs to balance the complexity and feasibility reasonably.

Design Alternatives and Final Approach

I. Alternative Solutions

At the beginning of the project, we extensively investigated and considered various modeling paths to realize an accurate prediction of the electrical performance of printed electronic devices. Our first candidate is the Convolutional Neural Network (CNN) image model. Based on the input of design drawings and printed images, a structure similar to pix2pix or image regression is constructed, which is suitable for the end-to-end process of "design drawings to predicted images to predicted electrical properties". The second candidate is the traditional machine learning method. Including linear regression, polynomial regression, LASSO regularization, multi-layer perceptron (MLP), etc., the direct mapping relationship of "parameters-performance" is established based on features, and the model structure is concise and interpretable.

II. Final Approach

Careful consideration of data characteristics and system objectives, as well as existing data. Finally, a traditional feature-based machine learning pipeline was ultimately adopted. In different data scenarios (R1 and R2), different traditional machine learning models are selected for modeling and comparison, and finally, the most suitable model is selected. In addition, we model the models with target and measured parameters as inputs, and make an in-depth evaluation through error comparison and an umbrella plot.

III. Rationale

This project finally adopts the decision of traditional regression modeling, mainly based on the following considerations:

i. Limited data size

The sample size of the project data set is small, and the data set is still given one after another with the progress of the project progresses. The existing data is not enough to support the effective training of the deep neural network model, and it is easy to overfit.

ii. The model requires high interpretability

In printing electronic technology, engineers need to clearly understand the logical relationship between "parameters-results". The explicit structure of linear and polynomial models is helpful for parameter sensitivity analysis and printing parameter adjustments, and can effectively correspond to related physical intuition.

iii. Limited computing resources and low deployment cost

Compared with an image class or a deep network model, the traditional model is more suitable for deployment in local low-computing environments, which is convenient for subsequent integration with printing equipment.

To sum up, it is more feasible and practical to adopt the characteristic engineering modeling method with a simple structure, stable performance and strong explanation under the current experimental conditions.

Design and Implementation

I. Data Collection Process

In order to establish the mapping relationship between geometric parameters and electrical properties, this project uses the DragonFly IV 3D printing platform of NanoDimension to print several groups of resistor samples with different geometric structures. The DragonFly IV 3D printing platform is shown in Figure 1.

FIG. 1. DragonFly IV 3D printing platform

The specific data acquisition process is as follows:

i. Sample design and printing

The printed object of this project is a two-dimensional resistor graph, and its main geometric parameters include line Width and Gap. The schematic diagram of the resistor structure is shown in Figure 2. In the design stage, we preset several target dimensions to cover the variation range of different widths and spacings.

On each printed board, we designed 9 groups of the same printed layouts, and each group of layouts contains several resistor patterns. The structural parameters in these layouts are set as follows. The line Width (width) starts from 150 μm and increases to 600 μm in steps of 50 μm. The Gap starts from 200 μm and increases to 425 μm in steps of 25 μm. A total of 10 × 10 = 100 parameter combinations are formed. In each set of layouts, there is a set of resistors on the left and right, respectively, and a total of 1800 resistors are printed. The actual image of the layout is shown in Figure 3.

In a complete print job, the printer will output 9 identical layout boards. Because each board contains 1800 resistors, a total of 16200 resistance samples are printed, which provides sufficient data support for subsequent model training and error analysis.

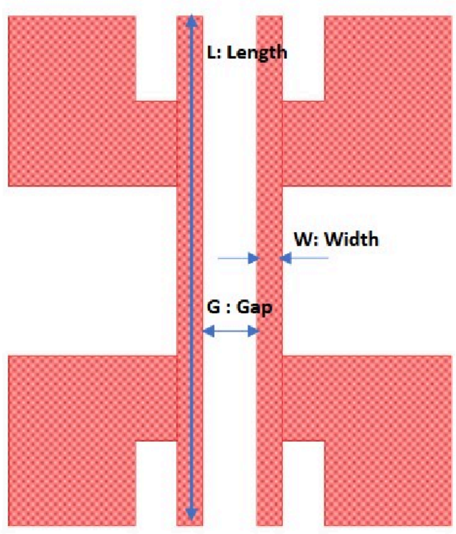


FIG. 2. Schematic of unit resistor pair showing length (L), width (W), and gap (G)

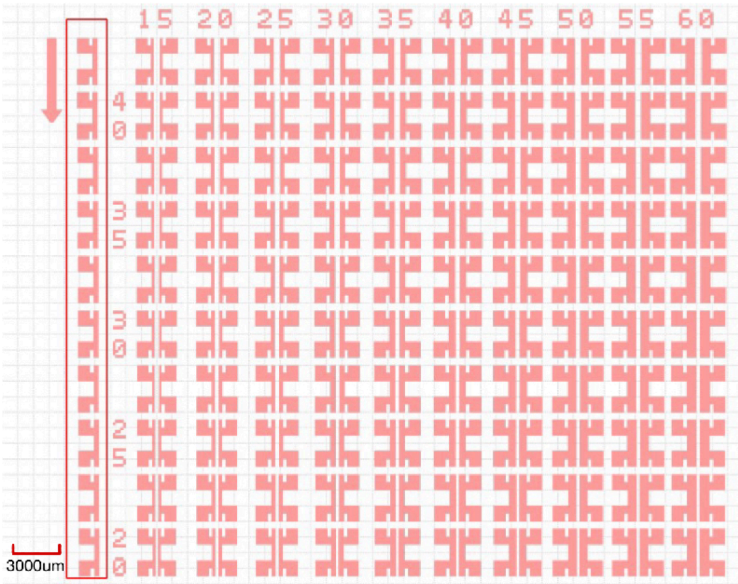


FIG. 3. Layout of printed resistor pairs for testing

ii. Geometric data recording

Target Dimensions: directly from the preset line width and spacing values in the layout design file.

Measured Dimensions: The samples are captured by a high-resolution optical microscope, and the line width and spacing are measured by image processing tools and recorded as the actual print dimensions.

In this way, each sample has two sets of inputs, the design value and the actual measured value, for subsequent comparative analysis and modeling.

iii. Electrical performance test

In order to measure the conductivity of the resistor, we use the Four-Probe Measurement to effectively eliminate the influence of contact resistance on the measurement results. We apply current to both ends of each sample, and measure the voltage through the two probes in the middle to ensure that it is a pure resistance voltage drop. The recorded current and voltage data are used to calculate resistivity. The measurement procedure is first, record both the target and measured widths/gaps. Then, perform four-probe measurements to obtain voltage and current, from which resistance and conductance are calculated. These data provide a complete and comparable input and output sample set for the subsequent machine learning model training. The actual measurement chart is shown in Figure 4.

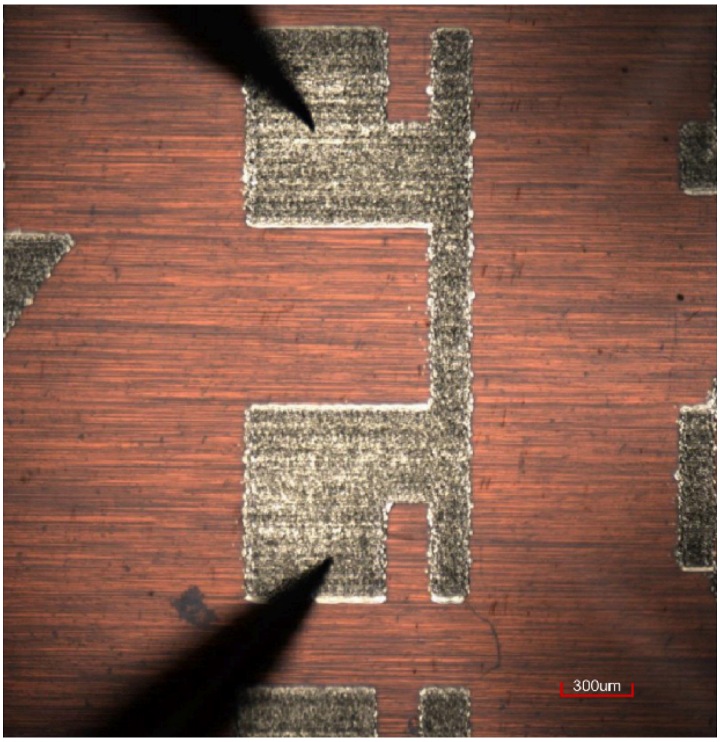


FIG. 4. Image of a physical printed resistor under four-probe measurement.

In the Four-Probe Measurement, we use a Keithley 2182A Nanovoltmeter, which is used for high-precision voltage measurement with a resolution of 1 nV and an input impedance of > 10 GΩ. Keithley 2450 SourceMeter: provides a constant low current (μA level) source and measures the current value at the same time. Both of them realize synchronous trigger through the 2450-TLINK trigger connection line, which constitutes the Delta mode measurement architecture. The resistor is connected in the middle of the system, and the current is introduced through two external probes, and the voltage is measured by two internal probes, thus effectively eliminating the influence of contact resistance. We use Delta Mode for testing. In Delta Mode, the source meter alternately outputs positive and negative currents, and the nanovoltmeter synchronously measures the corresponding voltages and takes the difference to offset the thermoelectric noise and system drift. This method is suitable for small voltage measurement of low resistance devices, and can significantly improve the signal-to-noise ratio and repeatability. According to Keithley's official technical specifications, the resolution of the system can reach 1 nV in the voltage range of 10 mV, the error is controlled within ±60 ppm (reading) within two years, and the temperature coefficient is less than 0.5 ppm/C, which meets the requirements of the project for high-precision measurement of micro-resistance devices.

II. Model Development

- **R1 Model (Using Width and Gap Only):**

i. R1 Model Setup

The R1 model aims to evaluate the influence of geometric parameters (line width and spacing) on the conductivity of **printed** electronic devices. We use two types of input features:

Target Dimensions: the original dimensions from the CAD layout design.

Measured Dimensions: the geometric dimensions of the finished product obtained by analyzing the microscope image.

The output variable is the Conductance of the device, and its physical meaning is the ease with which current flows in the structure.

ii. Motivations

The motivation of this part of the study mainly includes two aspects:

Firstly, explore the most suitable prediction model for this dataset, and compare the prediction accuracy among different models.

Secondly, the hypothesis of the linear relationship between conductivity and line width is verified, which can be deduced from the classical resistance formula $R = \rho \cdot l/w \cdot t$, which means that conductance should be proportional to width under certain conditions.

iii. Model Exploration

We tested the following three models to fit the relationship between conductance and target width/target gap and measured width/measured gap, respectively.

a) Linear Regression

First, we tried the linear regression model, which is also the most in line with our expectations.

Result of target dimension:

Conductance = 0.9314 + (0.0064×Target_Width) + (0.0009×Target_Gap); R^2 = 0.8844

The 3D plot of the linear regression results is as follows

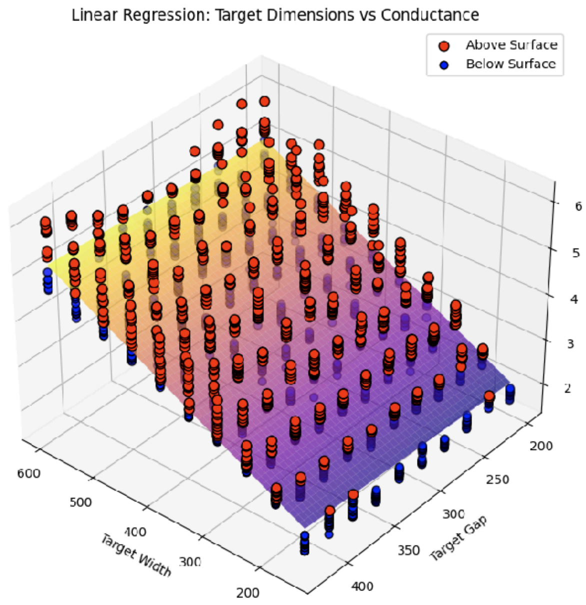


FIG. 5. Linear Regression 3D Plot of target feature

All the points are target values, which are evenly distributed. Red means to return to above the plane, and blue means below it. The image directly verifies the positive linear relationship between conductance, width and gap, which conforms to the physical law. Good prediction accuracy can be achieved by using target dimensions, which are suitable for the design stage with a simple structure. Although there are still some errors, the performance of the model is satisfactory without considering other process parameters.

The Slice plot of linear regression results is as follows:

FIG. 6. Target Linear Regression slices plot of 3D plot

The figure shows the comparison between the actual observed value (gray point) and the predicted value (red line) of conductance with Target Width under three subsets with fixed Target Gap values of 200um, 300um and 425um. In the low-width section (150-250 µm), individual samples fluctuate slightly, and the model prediction is slightly lower than the actual value, which may be related to the printing resolution error or the instability of the initial section. The fitting effect is the most stable in the middle width section (300-500 µm), and the actual data points are close to the prediction line. In the high-width section (550-600 µm), there is a slight systematic deviation of "the prediction is lower than the actual value", indicating that there may be a slight nonlinear upward trend. But the whole is still within the acceptable error range.

Result of measured dimension:

Conductance = 0.8705 + (0.0066 · Measured_Width) + (0.0009 · Measured_Gap); R^2: 0.88079

The 3D plot of linear regression results is as follows:

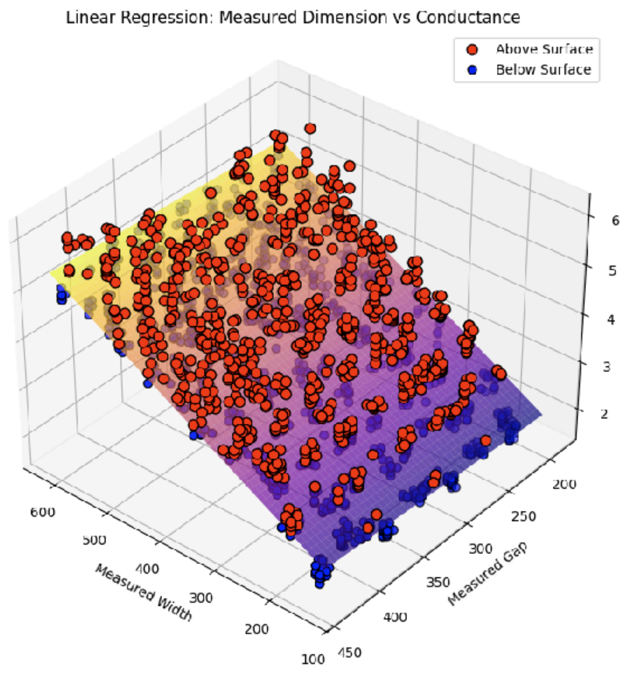


FIG. 7. Linear Regression 3D Plot of measured feature

All points in the graph are measured points; the color surface in the graph is a linear regression prediction surface, and the scattered points are actual data points. The Above Surface represents that the actual value is higher than the model prediction. The Below Surface indicates that the actual value is lower than the model prediction. The distribution of point clouds is close to the plane, which shows that the model has good fitting quality and no large-scale deviation.

The Slice plot of linear regression results is as follows:

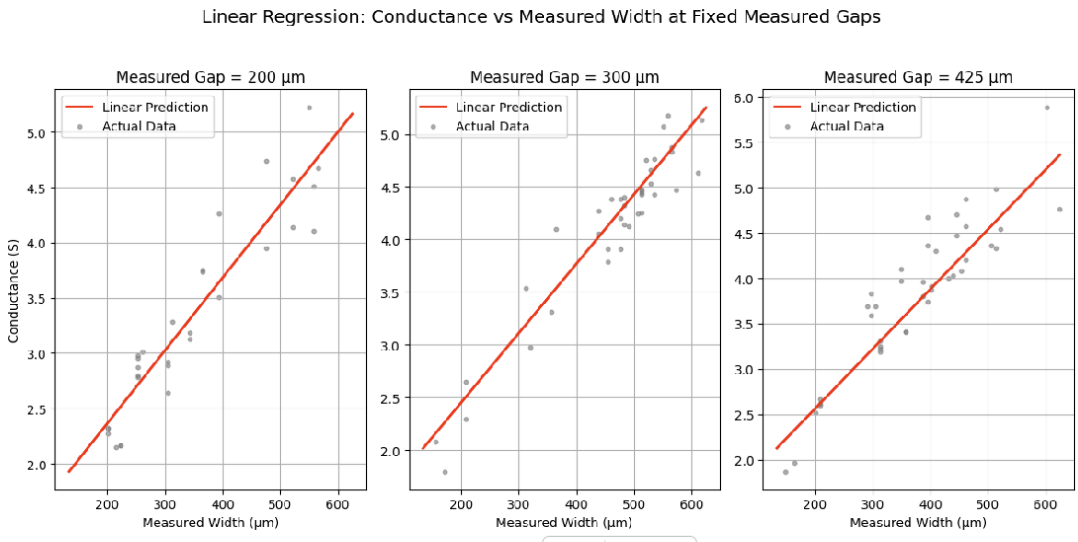


FIG. 8. Measured Linear Regression slices plot of 3D plot

The three subgraphs correspond to subsets with Measured Gap of 200 µm, 300 µm and 425 µm respectively. When Gap = 200 µm, the fitting performance is good, and only a few points deviate from the fitting line in the middle width section (250-450 µm), which may be caused by printing accuracy or measurement error. When Gap = 300 µm, the fitting effect is the best, and the height of actual data points is concentrated around the fitting line, which shows that the measured data under this gap condition is stable and the model fitting is the best. When Gap = 425 µm, some floating sample points appear in the local high-width section (500-600 µm), indicating that the actual value of conductance is slightly higher than the model prediction, which may indicate a slight nonlinear upward trend.

b) Multilayer Perceptron

Result of target dimension: R^2 score of the MLP model: 0.9056

The 3D plot of the MLP results is as follows

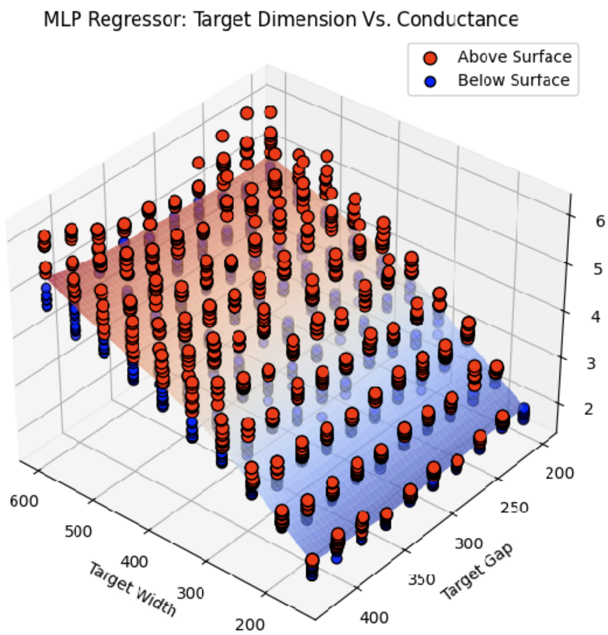


FIG. 9. MLP 3D Plot of target feature

Compared with the linear regression model, MLP is more expressive and can fit complex input-output relationships. As can be seen from the figure, the fitting surface is not completely flat, and there are slight bends and undulations. It shows that the model tries to capture the more complex interaction between width and gap and conductance. However, in the current data structure, this nonlinearity is not obvious and lacks significant physical support.

The Slice plot of the MLP results is as follows

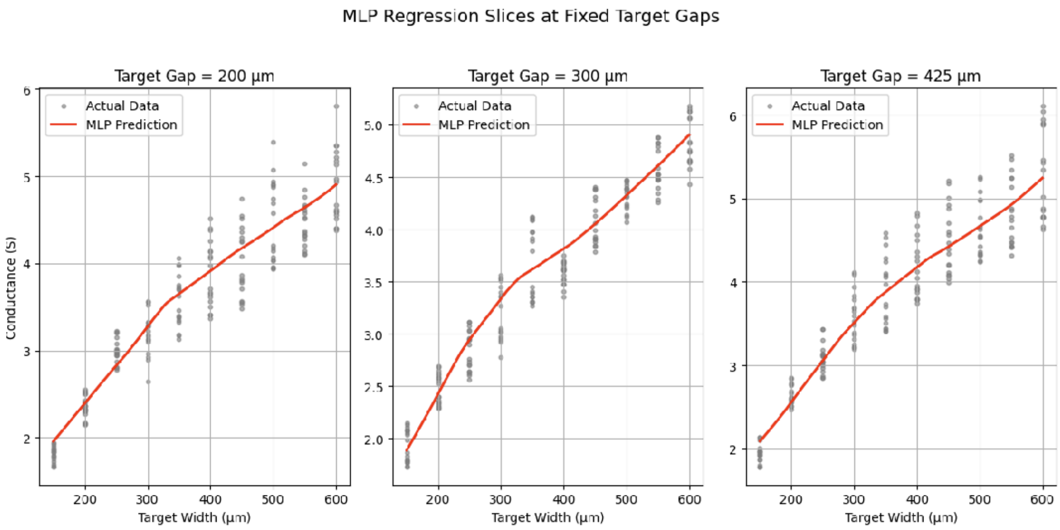


FIG. 10. Target MLP slices plot of 3D plot

Different from the straight line of linear regression prediction, the MLP fitting curve shows a slight nonlinear bending trend under the three gap conditions. Especially in the middle-width section (250-350 μm) and the high-width section (550-600 μm), the fitting line has some "bending". This may be that the edge effect or printing nonlinear error in the data is "captured" by the model.

Result of measured dimension: R^2 of the MLP model: 0.90054

The 3D plot of the MLP results is as follows

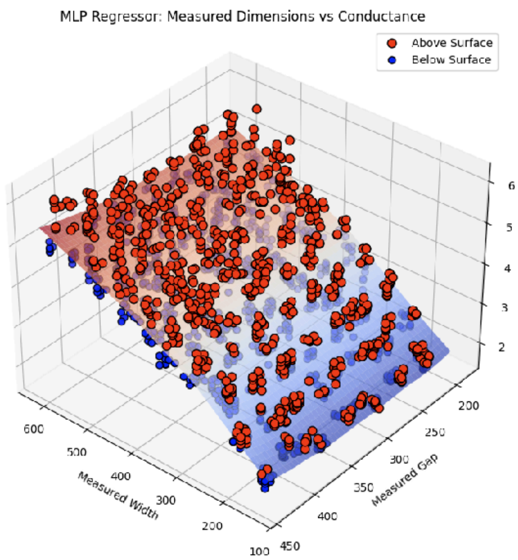


FIG. 11. MLP 3D Plot of measured feature

Most sample points are close to the fitting surface, which shows that the overall fitting quality of the MLP model is good. The model learns the nonlinear relationship between conductance and measured width/gap, especially in areas with small width or large gap. Compared with the target input model, the data points in this figure are denser, and the fitting shape is smoother, which effectively captures the subtle changes brought by the actual printing size.

The Slice plot of the MLP results is as follows

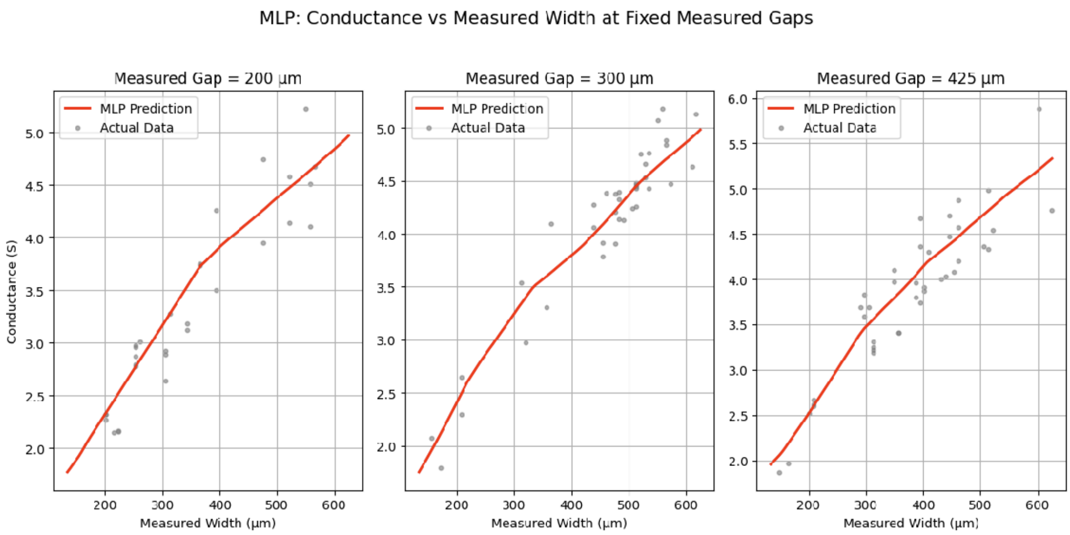


FIG. 12. Measured MLP slices plot of 3D plot

In the smaller width section (150-300 µm) and the larger width section (500-600 µm), the fitting line of MLP shows a smooth transition or a slight inflection point. In the middle section (300-500 µm), the fitting effect is highly consistent with the actual data, and the error distribution is small, but there is an obvious bending point. There are some outliers in the edge segment (low width/high width), especially at the right end of the gap = 425um and gap = 200um, and individual actual values deviate from the prediction.

c) Random Forest

Result of target dimension: R^2 of the Random Forest model: 0.9120

The 3D plot of the Random Forest results is as follows

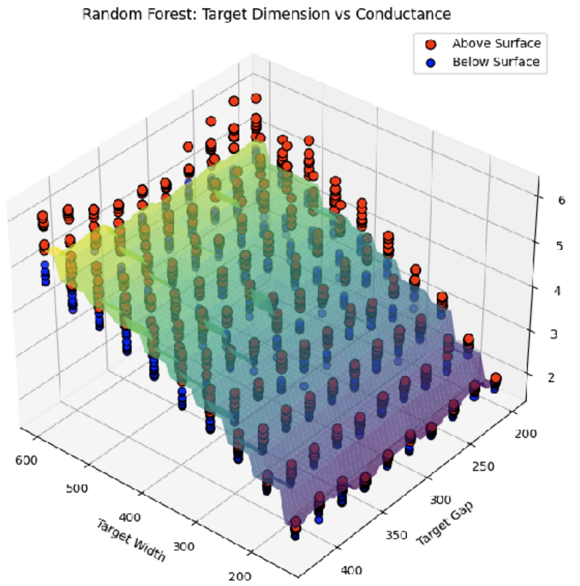


FIG. 13. Random Forest 3D Plot of target feature

Random forest constructs an integrated model composed of multiple decision trees, which can automatically capture complex characteristic nonlinearity and interaction items. The surface presents a relatively irregular "stepped" structure, which reflects the essence of piecewise constant prediction of the decision tree.

The Slice plot of Random Forest results is as follows:

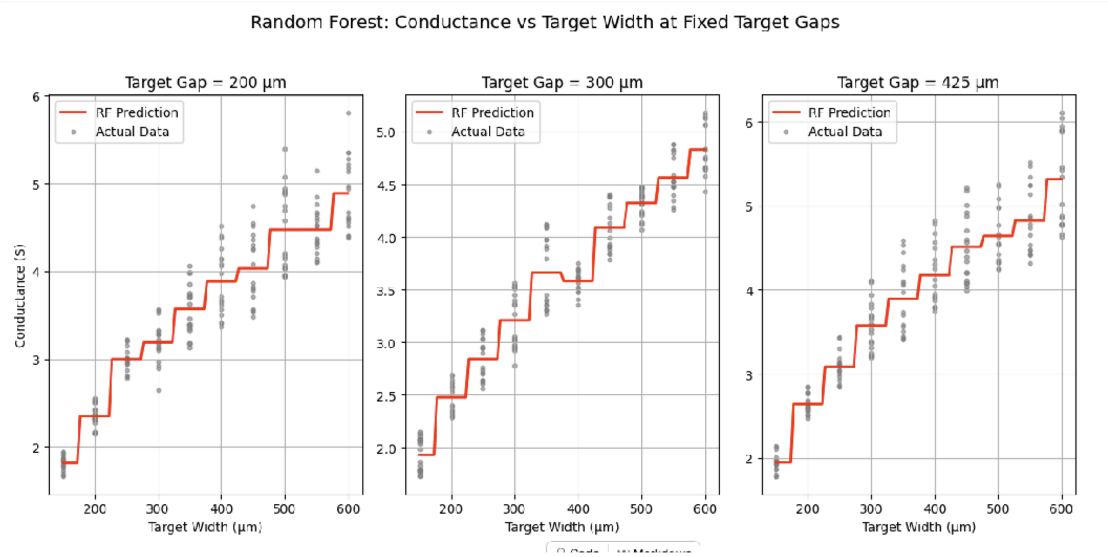


FIG. 14. Target Random Forest slices plot of 3D plot

The red prediction curve presents an obvious stepped discontinuous structure because the random forest model is essentially a piecewise constant model, which is composed of a large number of decision tree nodes. Each predicted value corresponds to the average value of a leaf node, so a smooth curve cannot be generated, but changes in a "jumping" way. We can't predict accurately, and at the same time, we can't get the input through the reverse of the results, so the model's effect is very poor.

Result of measured dimension: R^2 of the Random Forest model: 0.95112

The 3D plot of the Random Forest results is as follows

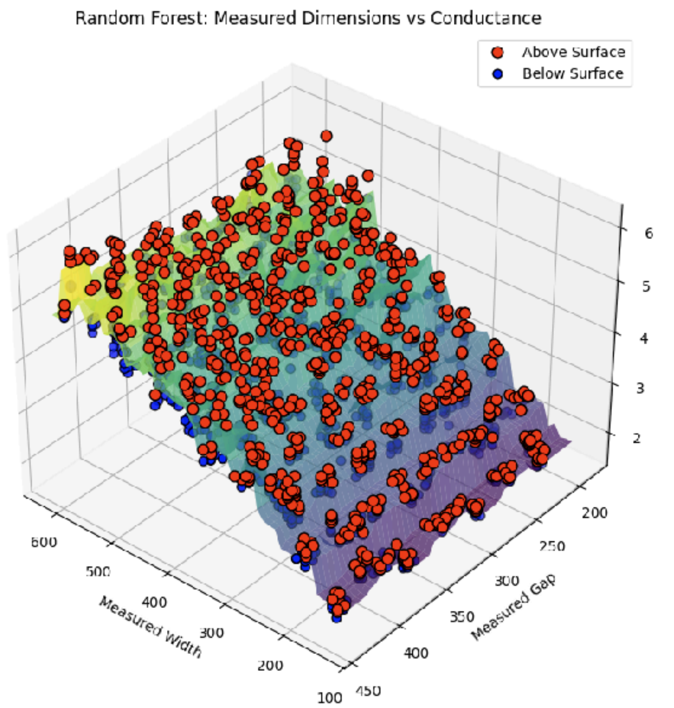


FIG. 15. Random Forest 3D Plot of measured feature

According to the image, it can be seen that the prediction surface is discontinuous, which is not suitable for physical modeling that requires high resolution/smooth output.

The Slice plot of Random Forest results is as follows:

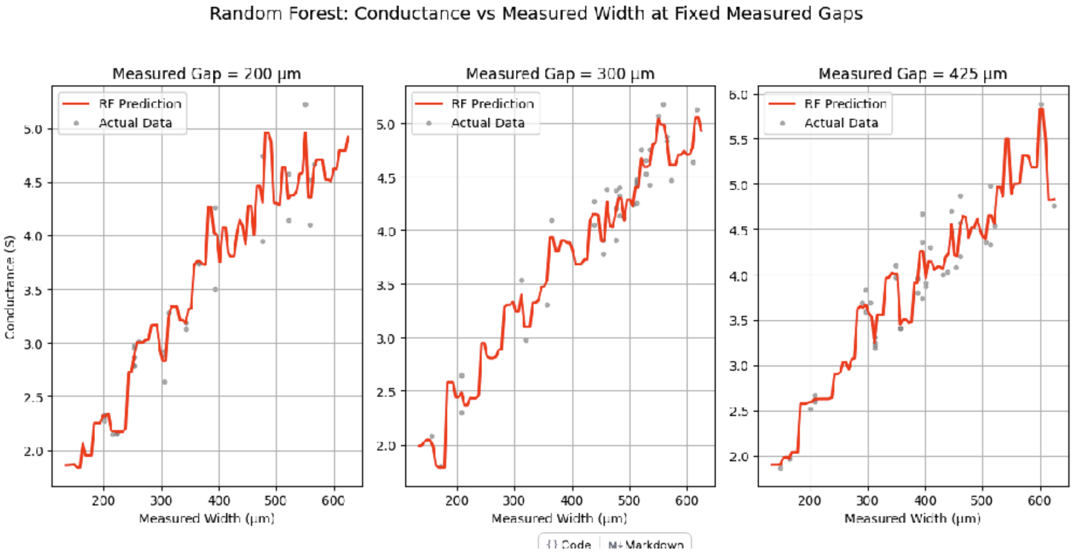


FIG. 16. Measured Random Forest slices plot of 3D plot

According to the slice diagram, we find that the Random Forest results of the measured dimension are obviously irregular. This shows that it is seriously over-fitted.

iv. Comparison

Based on the performance of various models and our project requirements, we finally chose the linear regression model as the core modeling scheme for conductivity prediction. The main reasons and comparison table is as follows:

a) Strong physical consistency: the structure of the linear regression model is consistent with our physical understanding of data and accords with the formula of the classical resistance model. Among them, conductance and width have a linear relationship. The model is concise and interpretable, which is convenient for subsequent analysis and system integration.
b) Although the MLP model is better in fitting, there is a risk of over-fitting: although MLP (Multilayer Perceptron) shows higher fitting accuracy in some indicators, the generalization ability and stability of the deep neural network are difficult to guarantee due to the limited dataset of this project. In addition, its complex black box structure is not conducive to the interpretation of results and the backtracking of physical parameters.
c) The Random Forest model does not meet the requirement of continuity of prediction: the prediction results of the Random Forest model are obviously stepped and discontinuous, which does not meet our requirements for smoothness and physical rationality of prediction values. Especially when using measured inputs, the model

has serious local oscillation and over-fitting behavior, which further reduces its practicality and credibility.

To sum up, linear regression is the best choice for the R1 dataset.

Model Name	**Performance (R1)**	**Do we use it**	**Why**
Multilayer Perceptron	0.9056(Target)/ 0.90054(Measured)	NO	Over fitting
Random Forest	0.9120(Target)/ 0.95112(Measured)	NO	Over fitting
Linear Regression	0.8844(Target)/ 0.88079(Measured)	YES	Met expectations

TAB. 1. Comparison between in models used in R1

To further compare the prediction error of the model when using the Target Dimensions and the Measured Dimensions as inputs, we draw the umbrella plot as shown in the figure. Based on linear regression, the figure presents the Percentage Error of the two models in the form of a vertical histogram. The blue part represents the model trained based on the target dimensions. The green part represents the model based on the measured dimensions training. The up-down direction in the figure does not represent the positive or negative error, but is only used to distinguish the model types, which is convenient for visual comparison of the range and density of error distribution. The umbrella plot is shown as follows:

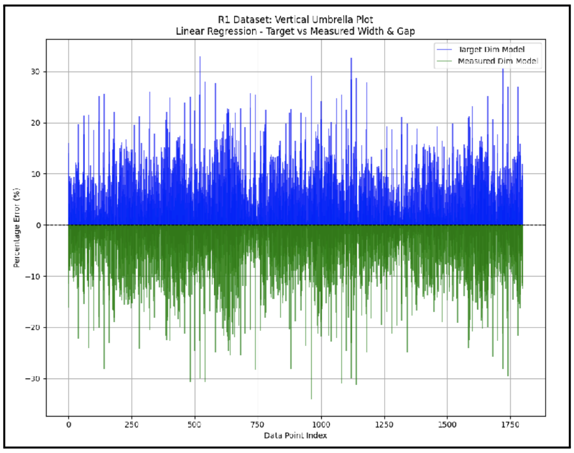


FIG. 17. R1 data Linear Regression Umbrella Plot

It can be observed from the figure that the error distribution range of the two models is equivalent, covering all sample points evenly. However, the error column of the measured model is "more compact" as a whole, and the error amplitude of the edge sample is slightly lower, which shows that it has better error stability and the ability to resist printing deviation. The error distribution of the target model is more sparse, and there are larger local errors in some samples. To sum up, this graph verifies the weak advantage of measured dimension input in prediction robustness and provides intuitive support for subsequent model selection.

- **R2 Model (Inclusion of Dimensions and Process Parameter):**

i. R2 Model Setup

With the physics of the printed resistor validated, the next phase of the project incorporates the printer's process parameters into the analysis. Specifically, it considers the Dynamic Top Oven Speed (DTO) and the Tray Temperature.

Dynamic Top Oven (DTO) Speed refers to the speed (mm/s) of the thermal curing element above the printed substrate. Adjusting this speed controls heat exposure during post-processing to reduce uneven ink flow. Among the settings, DTO 200 mm/s provides the most effective heating, followed by DTO 400 and DTO 0—the latter being cooler, while 400 mm/s is too fast for even heating. Tray Temperature (°C) indicates the platform's heat level, helping maintain substrate warmth, prevent thermal shock, and enhance uniform curing.

With these process conditions accounted for, the inputs to the R2 model include:

1. Target Width (µm): the expected width of the resistor as defined in the design layout
2. Target Gap (µm): the expected gap in the CAD layout
3. Measured Width (µm): the actual printed width, obtained through virtual metrology

4. Measured Gap (μm): the actual printed gap, also determined through virtual metrology
5. Dynamic Top Oven Speed (mm/s): 0, 200, 400
6. Tray Temperature (°C): 120, 160, 175

And the output of the model is the conductance of the printed resistor.

ii. Motivations

In the first stage (R1), the investigation focused solely on the relationship between the physical dimensions of the resistors and their electrical behavior. We confirmed that the theoretical foundation linking geometry to conductance was valid, as the observed data aligned with the expected linear correlation.

However, to fully evaluate the performance of the printed devices, it is essential to account for the printer's process parameters. Our approach targets the ink deposition and curing phases of the printing process, which are critical in determining the material properties and structural integrity of printed electronics. As such, these process parameters are treated as additional input variables.

By integrating this process information with virtual metrology data, the enhanced model (R2) allows for the isolation and analysis of individual process effects. This enables a more comprehensive assessment of the printer's predictability using a feature-based modeling approach.

iii. Model Exploration

In the second stage of the project, we evaluated four different models: Linear Regression, Multilayer Perceptron, Random Tree, and Polynomial Regression. Two input configurations were used for each model training. One combined target dimensions with the process parameters, while the other used measured dimensions from virtual metrology alongside the same process parameters. This allowed for a comparison between design-based and measurement-based predictions.

For each model, the correlation coefficient was calculated and the 3D plot of the prediction surface was plotted assuming constant gap and DTO values. Quantifiable models with optimal correlation coefficients will be further evaluated by scaling the coefficient to determine the weights of each variable. Note that the outliers of the dataset is filtered out before each round of training.

a) Linear regression

While the inclusion of additional process parameters disrupts the linearity previously observed in the R1 dataset, it provides valuable insight into how conductance dependencies evolve with more complex input interactions. Despite the loss of strict linear behavior, these models serve as a useful reference for understanding the deviations due to the added variables.

Result of target dimension:

Conductance = -7.9450 + (0.0050 * Target_Width) + (0.0003 * Target_Gap) + (0.0562 * Temp) + (0.0007 * DTO); R^2=0.7876

Result of measured dimension:

Conductance = −8.1205 + (0.0007 × DTO) + (0.0577×Temp) + (0.0050×Measured_Width) + (0.0001×Measured_Gap); R^2=0.79

The 3D plots of linear regression results are shown below. Gap effects were excluded due to their minimal impact on conductance and the limited dimensional range available for generating meaningful plots.

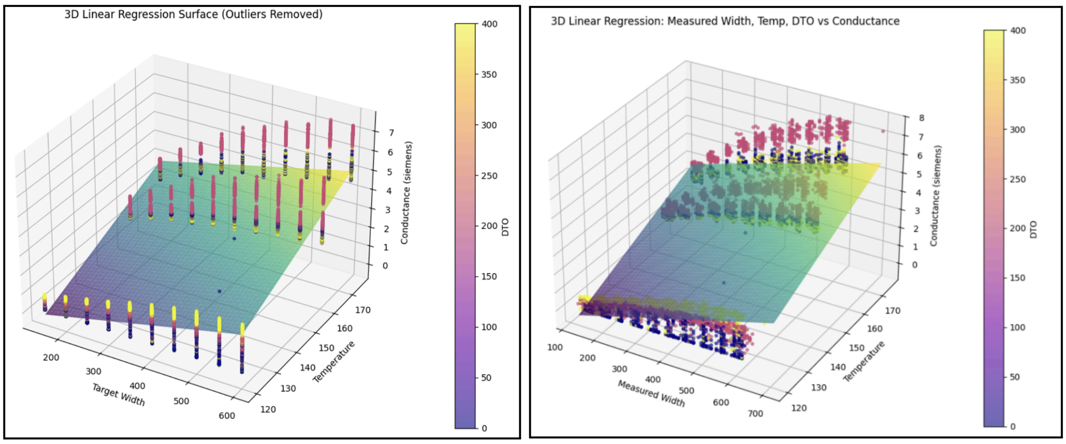


FIG. 18. 3D plots of linear regression using target dimensions (left) and measured dimensions (right) for the R2 dataset.

Overall, the correlation coefficient for both models is similar, with the measured dimension slightly better. According to the plot, the dataset appears in vertically aligned columns and is randomized in the two plots, corresponding to the fixed intervals of the layout and natural fluctuations in the actual measurements. Color-mapped DTO values reveal a consistent trend: at 120 °C, 400 mm/s, DTO produces higher conductance, and at 0 mm/s, the lowest. This trend reverses at 160–170 °C, indicating a nonlinear, temperature-dependent relationship. The regression surfaces show general trends but fail to capture the curvature and edge effects, particularly where conductance nears zero at 120 °C, indicating limitations in modeling nonlinear behaviors.

To examine the effects in greater detail, 2D slices at fixed temperatures and mean DTO values are shown below.

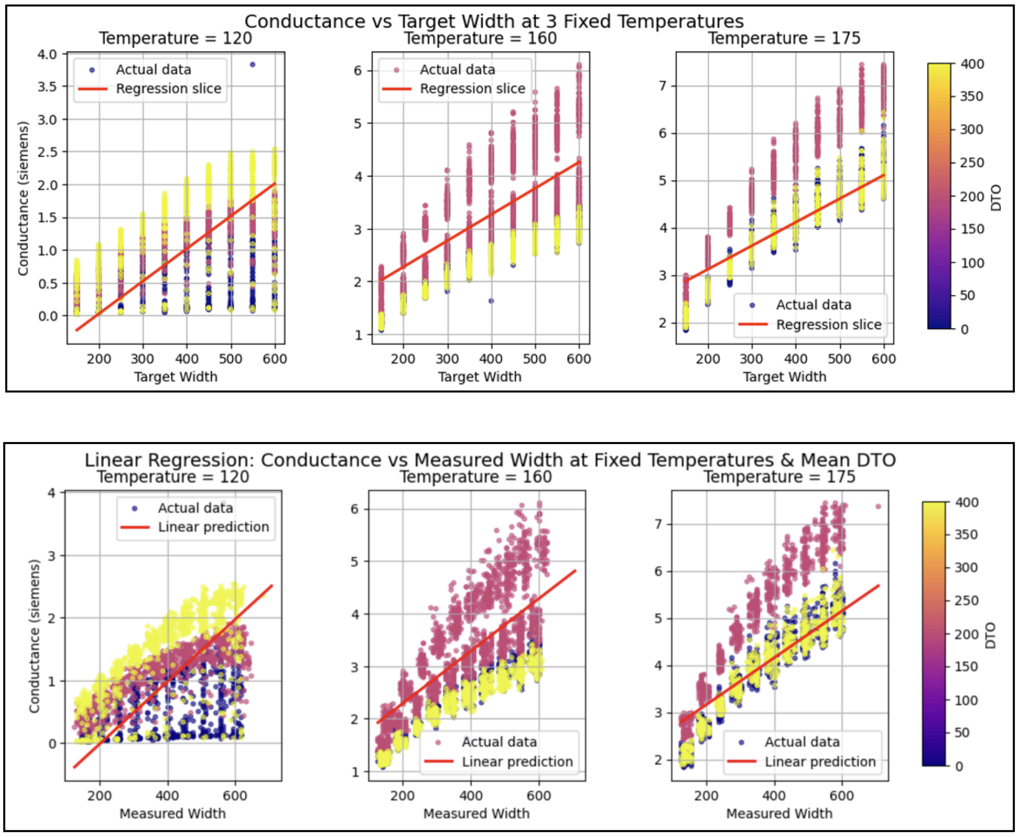


FIG. 19. Linear regression slices from the 3D plots using target dimensions (top) and measured dimensions (bottom) at three fixed temperatures and mean DTO.

These 2D views highlight fitting accuracy across temperatures. At 120 °C, regression curves are notably pulled downward due to low conductance values at 0 mm/s DTO, with some near zero, resulting in an overestimation of the fitting line. At 160 °C and 175 °C, while the fit improves, visible nonlinear patterns of the dataset persist, especially in the measured width plots. These deviations emphasize that a simple linear model struggles to fully capture the temperature- and DTO-dependent behavior of conductance.

b) Multilayer Perceptron (MLP)

This model was expected to predict better than linear regression does, since the nature of the model allows capturing nonlinearity in the measurement data. Nevertheless, there is no analytical form of the trained result, and the fitting surface can only be extracted by a numerical approach. The correlation coefficients of the two trials are presented below.

Result of target dimension: R^2=0.825

Result of measured dimension: R^2=0.7603

The 3D plots of training results using MLP are shown below. Gap effects were excluded.

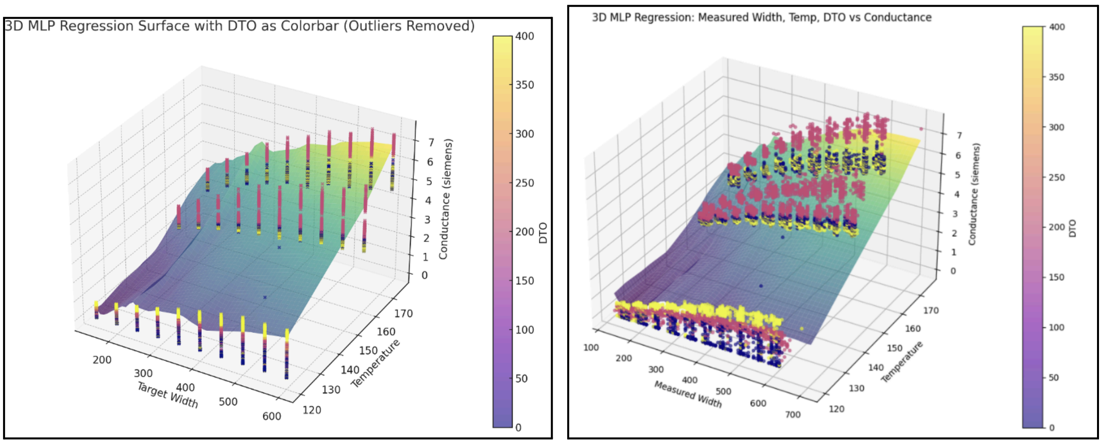
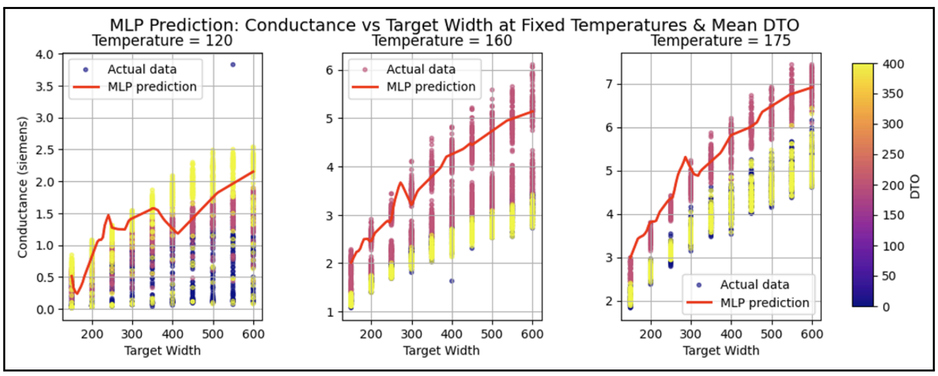


FIG. 20. 3D plots of Multilayer Perceptron using target dimensions (left) and measured dimensions (right) for the R2 dataset.

Based on the correlation coefficients, the target model outperforms the measured model, likely due to the greater fluctuations present in the measured dataset. From the plots, the target and measured datasets exhibit identical behavior to that described in the linear regression discussion. The fitting curve, however, introduces some curvature to account for the impact of additional DTO and temperature variables. The surface fluctuates most at smaller target/measured widths below 300 μm, and smooths out at larger widths (above 500 μm for target dimensions and 400 μm for measured dimensions).

Similarly, slices of the fitting curve at fixed temperatures and mean DTO are shown below:



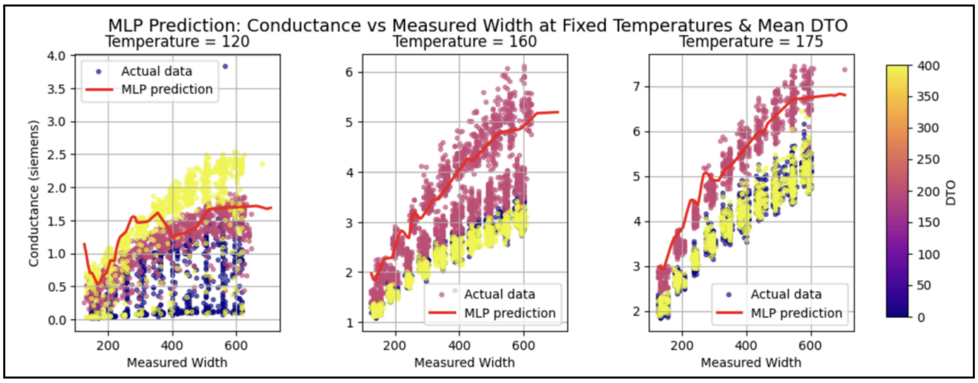


FIG. 21. Multilayer Perceptron slices from the 3D plots using target dimensions (top) and measured dimensions (bottom) at three fixed temperatures and mean DTO.

From the slices, there is significant misalignment at 120 °C. At this temperature, the fitting curves trained using both target and measured dimensions fluctuate irregularly around the upper portion of the dataset, exhibiting erratic up-and-down trends that do not meaningfully reflect the underlying data distribution. This inconsistency reduces above 400 µm, where the curves begin to smooth out and align more closely with the central trend of the conductance values.

In contrast, the slices at 160 °C and 175 °C show much better alignment, particularly at 160 °C. The fitting curves in these cases tend to track the main clusters of data across various width and gap combinations without displaying unnatural fluctuations. However, at 175 °C, some signs of overfitting emerge, with the curve abruptly adjusting its direction to follow individual data points in a stepwise manner.

c) Random Forest

Since the Random Forest model represents the dataset using discretized steps, it is expected to be prone to errors in reverse prediction (i.e., using conductance to infer design dimensions). Nonetheless, we aim to explore how this imperfect model might inform and inspire future testing strategies. Like the MLP model, Random Forest lacks an analytical form. The corresponding statistics and plots are presented below:

Result of target dimension: R^2=0.9519

Result of measured dimension: R^2=0.9893

The 3D plots of training results using Random Forest are shown below. Gap effects were excluded.

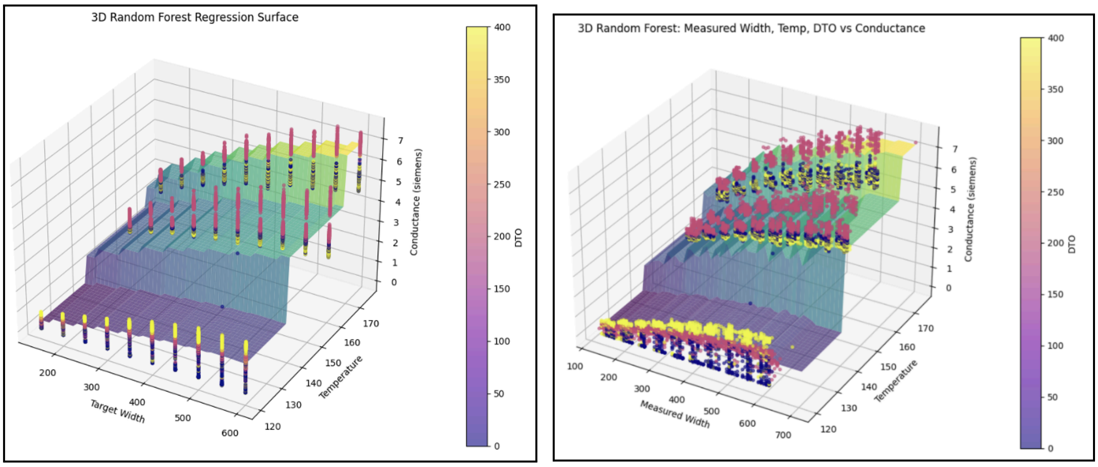


FIG. 22. 3D plots of Random Forest using target dimensions (left) and measured dimensions (right) for the R2 dataset.

The correlation coefficient appears excellent, approaching one for the model trained on measured dimensions. However, this model is ultimately unsuitable for adoption for two key reasons. First, the high correlation is a result of overfitting; the fitting surface is discretized to closely match specific segments of the dataset, leading to pronounced fluctuations and instability in the prediction curves. Second, the abrupt transitions in the surface make it impractical for reverse prediction, as even minor differences in desired conductance can correspond to widely varying design values. For these reasons, this method is fully discarded from consideration.

d) Polynomial Regression

Based on previous testing of the dataset's trends, we expected Polynomial Regression to yield the optimal results, as the data appeared to follow a distribution that a polynomial curve could effectively capture. The analytical expressions of the trained second-order Polynomial Regression are shown below:

Result of target dimension:

Conductance = 2.8059 + (0.7041 * Target_Width) + (0.0128 * Target_Gap) + (1.5406 * Temp) + (0.0904 * DTO) - (0.1281 * Target_Width^2) + (0.2565 * Target_Width Temp) + (0.0207 * Target_Width DTO) + (0.0097 * Target_Gap Temp) + (0.5242 * Temp^2) - (0.1569 * Temp DTO) - (0.4830 * DTO^2); R^2=0.9214

Result of measured dimension:

Coductance = 2.8017 + (0.7048 * Measure_Width) + (0.0123 * Measure_Gap) + (1.5678 * Temp) + (0.0914 * DTO) - (0.1287 * Measure_Width^2) + (0.2490 * Measure_Width Temp) + (0.0212 * Measure_Width DTO) + (0.5304 * Temp^2) - (0.1570 * Temp DTO) - (0.4727 *

DTO^2); R²=0.922

It is worth noting that the expressions above are the result of post-processing using the Least Absolute Shrinkage and Selection Operator (LASSO), as the original coefficients alone do not reveal whether a variable is truly essential or simply has a small magnitude that would require significant scaling.

The 3D plots of training results using Random Forest are shown below. Gap effects were excluded.

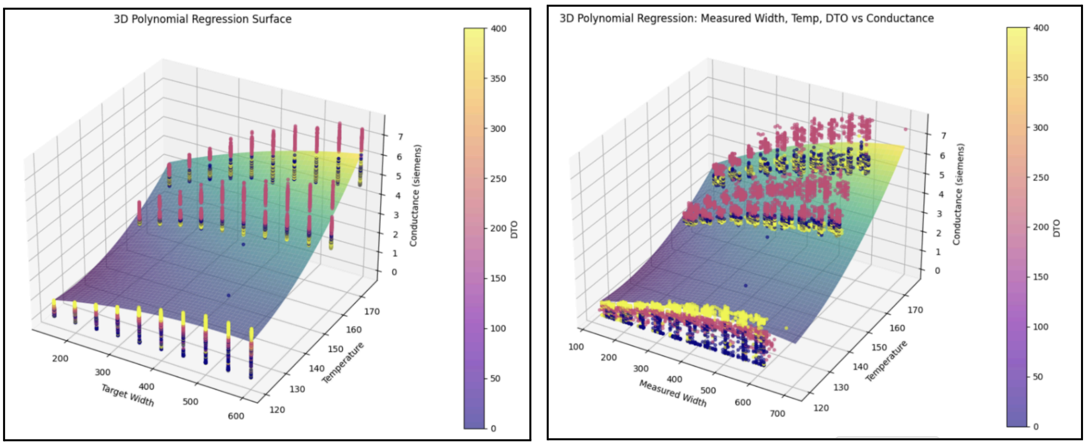
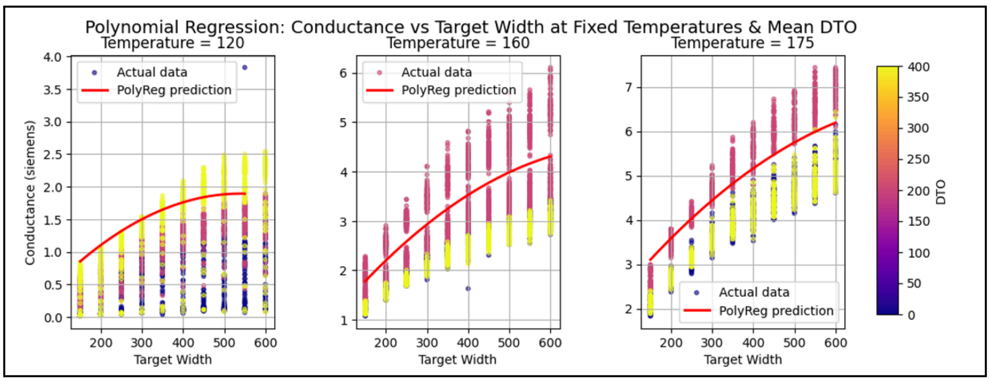


FIG. 23. 3D plots of Polynomial Regression using target dimensions (left) and measured dimensions (right) for the R2 dataset.

The correlation coefficients of the two trials are both slightly above 0.9, potentially indicating a strong fit. From the 3D plots, the fitted surface demonstrates reasonable coherence with the data distribution. The curvature of the surface aligns well with the variation in peak values across the data columns. A closer look at the slices at fixed temperatures and mean DTO is presented below:



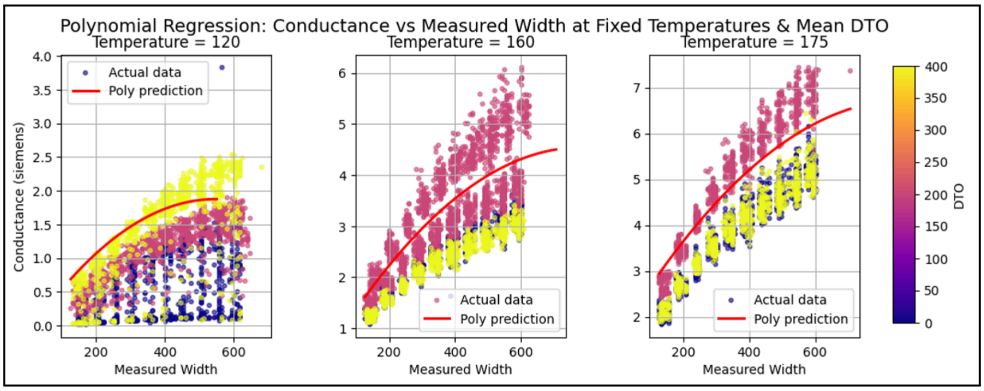


FIG. 24. Polynomial Regression slices from the 3D plots using target dimensions (top) and measured dimensions (bottom) at three fixed temperatures and mean DTO.

Importantly, portions of the polynomial slices showing downward trends were not plotted due to their lack of meaningful interpretation. From the slices, overfitting issues persist at 120 °C due to the scarcity of data points. Still, the nonlinear increasing trend of conductance is captured by the prediction curve with reasonable accuracy. Specifically, at higher temperatures, the distribution's characteristics are well modeled, with the polynomial approximately dividing the dataset in half. A smooth and continuous pattern is also observed, supporting potential extrapolation of conductance behavior. Overall, the polynomial regression model demonstrates acceptable accuracy and correlation, making it a viable option for prediction among our testing results.

Model Evaluation

An umbrella plot was also generated for the R2 model, as shown below. The plot exhibits near symmetry across the axes, indicating that the prediction errors are comparable whether target width or measured width is used as input. Given this similarity in performance, and for the sake of convenience in future model development, we will proceed using the target width.

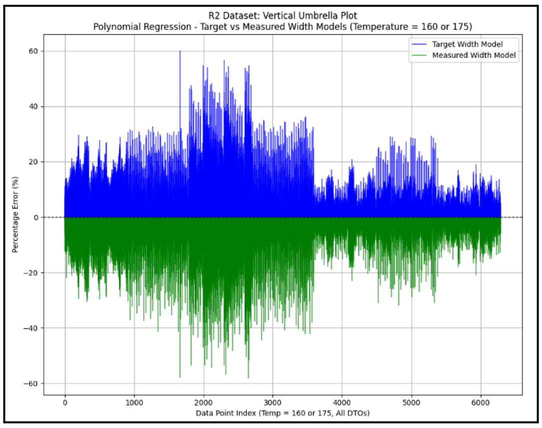


FIG. 25. Umbrella plot showing the error distribution between training inputs using the target dimension (blue) and the measured dimension (green).

Results

The model development for this project was conducted in two phases—R1 and R2—each targeting the prediction of conductance in 3D-printed electronics using progressively more complex input features.

I. **R1 Model (Geometry-Based Prediction Only):**
The R1 model focused solely on geometric parameters—line width and gap—using both target (design-intended) and measured (post-print) dimensions as inputs. Three machine learning models were evaluated: Linear Regression, Multilayer Perceptron (MLP), and Random Forest.

i. Linear Regression yielded strong performance ($R^2 \approx 0.88$) and aligned well with the physical theory of conductance, showing clear linear trends between width and conductance.

ii. MLP slightly improved the fit ($R^2 \approx 0.90$) and captured mild nonlinear behaviors, but at the cost of interpretability and a higher risk of overfitting given the small dataset.

iii. Random Forest achieved the highest R^2 (up to 0.95 with measured dimensions), but produced discontinuous predictions and exhibited strong signs of overfitting.

Ultimately, Linear Regression was selected as the R1 model due to its physical consistency, interpretability, and robustness, despite slightly lower raw performance metrics. The umbrella plot analysis supported this choice by showing tighter error distributions with measured inputs, indicating better robustness to real-world deviations.

II. **R2 Model (Geometry + Process Parameters):**
The R2 model expanded on R1 by incorporating process variables: Dynamic Top Oven (DTO) speed and tray temperature. The model evaluation included Linear Regression, MLP, Random Forest, and Polynomial Regression (with LASSO regularization).

i. Linear Regression performed reasonably ($R^2 \approx 0.79$) but struggled to capture the nonlinear effects introduced by process parameters.

ii. MLP slightly improved the fit for target dimensions ($R^2 = 0.825$) but underperformed with measured inputs, indicating higher sensitivity to data noise.

iii. Random Forest showed very high R^2 values (up to 0.989), but its discontinuous, step-like prediction surface made it unsuitable for physical modeling and reverse engineering.

iv. Polynomial Regression with LASSO emerged as the most balanced solution ($R^2 \approx 0.92$), capturing nonlinear interactions while maintaining smoothness and interpretability in its analytical form.

Given the performance and structure, Polynomial Regression with LASSO was chosen as the R2 model. It provided the best trade-off between fitting power, physical realism, and robustness. Slice plots showed that the model generalized well at higher temperatures and offered a continuous prediction surface aligned with process effects.

In conclusion, the project successfully developed interpretable and accurate machine learning models for predicting conductance in printed electronics, progressively incorporating more physical and process-based variables. Linear regression served well for simpler geometrical analysis, while polynomial regression with LASSO became the preferred solution under more complex, multivariable conditions. The findings not only validate core physical hypotheses but also lay the foundation for creating a data-driven “digital twin” for rapid prototyping and optimization in printed electronics manufacturing.

What Worked / What Didn’t

I. What Worked

This project focuses on the mapping relationship between geometric parameters and the conductivity of printed electronic devices, and builds and compares different prediction models based on two types of printed data sets. At the same time, we systematically discuss the influence of the difference between the target dimensions and the measured dimensions as the model input on the prediction performance.

The modeling process of this project has achieved good results in many aspects. Firstly, according to the data characteristics of different stages, we flexibly select and train the model with the strongest adaptability. In the R1 data set with geometric parameters, the linear regression model not only has high fitting accuracy but also has strong physical interpretability. However, in the R2 data set with process parameters, the polynomial regression model of Lasso regularization balances the prediction performance and generalization ability better. They all demonstrated a strong correlation without overfitting. Secondly, we systematically compare the influence of design size (target) and actual measured size as input on the model performance, and clearly show the error structure through umbrella

plot and other visualization methods.
Finally, the machine learning models trained in the whole modeling process show strong input-output correlation, and relatively low prediction errors are obtained in both stages, which verifies the rationality of our feature selection, model selection and evaluation methods.
Model matching strategy based on data structure differences, feature selection guided by physical intuition, and input optimization driven by error comparison analysis all constitute the core design elements of our prediction system to achieve good performance.

II. What Didn’t Work

Although the overall modeling effect of this project is good, there are still some factors that affect the performance. Firstly, the measurement error and printing error in the input data set have a negative impact on the model's performance. Due to the deviation or instability of some measured dimensions in the printing process, and the inaccuracy of electrical measurement results. It leads to uncontrollable noise when training the model, which further affects the prediction accuracy, especially under the edge conditions such as high width or large spacing.

Secondly, the scale of the training data set is limited, which limits the complexity selection of our model to some extent. For example, although Multilayer Perceptron (MLP) and random forest have stronger nonlinear fitting ability in theory, they are more prone to overfitting under small sample data, and the generalization performance is difficult to guarantee. In addition, the small amount of data also limits our in-depth modeling of the interaction between some input features (such as DTO or temperature).

Therefore, although we compensate by means of feature selection and regularization, the robustness and expansibility of the model are still restricted. If the data volume can be further improved and the measurement error can be reduced in the future, there is still room for significant improvement in model performance.

Future Work

On the basis of this project, future research can be further expanded and optimized in the following directions.

I. Expanding data scale and improving data quality
The current model is limited by the volume and measurement accuracy of data sets. In the

future, the generalization ability and prediction accuracy of the model can be further improved by collecting larger and more accurate printed sample data, especially under the conditions of atypical structure and boundary parameters.

II. Expand the research scope and circuit types

At present, the modeling object is mainly a simple resistor structure, and more types of layout can be designed later, including the printed structure of capacitors, inductors and other components, thus expanding the applicability of the prediction framework in more complex circuit modules. In addition, more printing process-related parameters, such as environmental humidity, ink type and nozzle structure parameters, can be introduced to describe the manufacturing process more comprehensively.

III. Try more advanced machine learning methods

In addition to the existing models, we can explore more expressive model structures, such as Gradient Boosting and support vector regression (SVR), to further improve the modeling ability of complex nonlinear relationships.

IV. Image-based Deep Learning Modeling

The actual layout or mask image is introduced into the model input, and the image-to-electrical performance prediction model is constructed by using technologies such as a convolutional neural network (CNN), which is expected to establish an end-to-end model from layout design to performance prediction and improve the automation design capability.

V. Develop a platform for real-time prediction and parameter optimization.

Based on the existing model, interactive front-end interface or lightweight simulation tools can be further developed to realize real-time electrical performance prediction and parameter adjustment suggestions after printing parameters are input, providing engineers with a more intuitive and efficient design auxiliary platform.

Conclusion

This project successfully developed and evaluated feature-based machine learning models to predict the electrical properties of 3D-printed resistors based on geometric and process parameters. The resulting models demonstrated strong predictive performance within the current experimental scope, though the size and variability of the available dataset limit their applicability. Overall, the work contributes to improving the predictability and efficiency of the FHE printing process, offering a data-driven approach to reduce material waste and optimize printing accuracy.

Appendices

Appendix A: Sample Dataset (CSV Format)

A sample of the dataset used for model training and evaluation is provided as a .csv file in this appendix. The file contains both input features and target outputs used for the R1 and R2 models.

File Name: training_data_sample.csv

Column Descriptions:

Column Name	Description
Target_Width	Width of the resistor as designed in the lithography layout (in microns)
Target_Gap	Gap of the resistor as designed (in microns)
Measured_Width	Actual printed width of the resistor (in microns)
Measured_Gap	Actual printed gap of the resistor (in microns)
DTO_Speed	Speed of the Dynamic Top Oven (DTO) during printing (in mm/s)
Temperature	Temperature of the top oven during printing (in °C)
Conductance	Measured electrical conductance of the printed resistor (in Siemens)