An Automated Machine Learning Approach to Inkjet Printed Component Analysis: A Step Toward Smart Additive Manufacturing

Abhishek Sahu¹, Peter H. Aaen² and Praveen Damcharla³
¹Qorvo Us Inc., Greensboro, NC, USA
²Colorado School of Mines, Golden, CO, USA
³KineticAI Inc., The Woodlands, TX, USA
¹abhishek.sahu@ieee.org, ²paaen@mines.edu, ³praveen@kineticai.com

Abstract—In this paper, we present a machine learning based architecture for microwave characterization of inkjet printed components on flexible substrates. Our proposed architecture uses several machine learning algorithms and automatically selects the best algorithm to extract the material parameters (ink conductivity and dielectric properties) from on-wafer measurements. Initially, the mutual dependence between material parameters of the inkjet printed coplanar waveguides (CPWs) and EM-simulated propagation constants is utilized to train the machine learning models. Next, these machine learning models along with measured propagation constants are used to extract the ink conductivity and dielectric properties of the test prototypes. To demonstrate the applicability of our proposed approach, we compare and contrast four heuristic based machine learning models. It is shown that eXtreme Gradient Boosted Trees Regressor (XGB) and Light Gradient Boosting (LGB) algorithms perform best for the characterization problem under study.

Index Terms—AutoML, Inkjet printing, LightGBM, Printed electronics, ResNet, XGBoost

I. INTRODUCTION

Printed electronics has gathered wide attention in the emerging markets such as internet of things as it offers biodegradable and cost-effective solutions. Recently, many traditional methods such as gravure printing, flexography, screen printing, and inkjet printing have been introduced in manufacturing of microwave circuits [1]. Among them, inkjet printing is most popular in the scientific and industrial community due to its attractive features including low manufacturing cost, large-area processability and lower carbon footprints. RF/microwave community has reported numerous applications of inkjet printing including antennas, wireless power transfer topologies, sensors, and microwave components [2]–[5] for low-loss and high-speed communication systems. Significant effort has been devoted from the research community to extract the material parameters of inkjet printed components from their measurements [6]–[10]. Conventional approaches need high precision profile measurements to extract the ink conductivity and a conformal mapping method (CMM) based on elliptical integrals to evaluate the dielectric constants. However, as these analytical methods are manually calculated to extract the parameters independently, they require rigorous human effort. Recently, computer-aided design (CAD) based-techniques utilizing commercial EM simulation packages and measurements have been proposed for automation and concurrent parameter extraction [10]–[13]. The problem with this approach is it requires a large number of iterative simulations. Further, as the number of variables increase, the optimization space becomes very complex and sometimes may lead to convergence problems. Therefore, there is a need for an alternative characterization approach that offers a more straightforward automation and is able to extract multiple parameters simultaneously without increasing process complexity.

Recently machine learning based approaches have gained wide attention as they make the characterization problem computationally intelligent [14], [15]. However, machine learning based approaches require significant expertise. Additionally, extensive human effort is needed to select an appropriate algorithm and their hyper parameters for each problem. To overcome aforementioned problems, we
present an automated machine learning models (AutoML) for accurate extraction of the material parameters of inkjet printed components. The proposed approach is fast and robust and is able to obtain all of the material properties with only a few sample of measured data.

The paper is organized as follows. Section II presents the detailed procedure to extract the propagation constant from measured S-parameters and a multiline TRL calibration. Section III presents the details of the machine learning models. Section IV presents the material characterization using machine learning models and compares the performance of the proposed models. We finally conclude the paper in section V.

II. DATA EXTRACTION

A. Fabrication and Measurement

To demonstrate the applicability of our proposed approach, we fabricated a set of CPW lines using a commercial process with Inkjetflex [16]. The details of fabrication can be found in [11]. The CPW lines (Fig. 1) have a 1.983 mm center conductor (W_c), a 0.13 mm gap (g), and 1.983 mm wide ground planes (W_g) fabricated on 125-μm-thick (t_FS) PET substrate with approximately 2 μm copper plating. To complete a benchmark multiline TRL calibration, the fabrication included a set of 6 CPWs with lengths of l=14.97 mm, 18.42 mm, 23.57 mm, 35.78 mm, 61.60 mm, 97.93 mm), a symmetric short circuit and a reflect standard. The CPWs were supported by a 0.5-cm-thick (t_DS) flexible substrate to print CPWs with a ground plane of width (W_c), center conductor of width (W_g) and gap of width (g). The CPWs are supported by a dielectric spacer of thickness (t_FS) underneath.

Fig. 1 shows the measurement setup for the inkjet printed CPWs. An Agilent E8364B vector network analyzer and an on-wafer probing system were used to carry out S-parameter measurements from 10 MHz to 20 GHz. The probing system includes two ground-signal-ground microwave probes (1000-μm pitch), probe manipulators, and an optical microscope. The power level of VNA was set to -17.0 dBm. An HP 4285A LCR meter was connected with the VNA to measure the dc resistance of the THRU. During measurements, the CPWs were supported by a 0.5-cm-thick (t_DS) porous, dielectric spacer to mitigate any potential deflection of the flexible PET substrate and avoid propagation of any undesired parallel-plate and microstrip modes due to the metallic vacuum chuck.

We employed a multiline TRL calibration to extract the complex propagation constant from uncalibrated S-parameter measurements. It should be noted that, in multiline TRL, the calibration is defined relative to the characteristic impedance of the transmission line (Z_0). Additionally, the propagation constant is extracted by analytically solving the eigenvalues of the measured cascade matrices as [17]

\[
\lambda_{ij}^{1,2} = \frac{1}{2} \left[ (M_{ij}^{11} + M_{ij}^{22}) \pm \sqrt{(M_{ij}^{11} - M_{ij}^{22})^2 + 4M_{ij}^{12}M_{ij}^{21}} \right]
\]

where \(M_{ij}^{ij}\) represent the measured cascade matrices of line pair i and j. For a two port device, the relationship between S and M parameters can be obtained as

\[
M = \frac{1}{S_{21}} \begin{bmatrix} (S_{12}S_{21} - S_{11}S_{22}) & S_{11} \\ -S_{22} & 1 \end{bmatrix}.
\]

The accuracy of the multiline TRL calibration can be validated by a calibration comparison method as described in [17].

B. EM Modeling

In order to obtain the training data for the machine learning models, full-wave finite element EM simulations were performed for the inkjet printed CPWs. Fig. 2 shows the cross section of the CPWs in an EM simulation environment. The full-wave EM simulation accounts for all terms of Maxwell’s equations and thus takes into account radiation losses in α. Single mode propagation is crucial for multiline TRL calibration. Therefore, to emulate the measurement setup, we ensured that the propagating EM field in the finite element model contains only one mode and no higher order modes are present in the waveguide ports. Once the EM simulation set up is validated, parametric analysis was performed by sweeping each of the input parameters (σ_ink, ε_FS, ε_DS, tanδ) to obtain the corresponding output parameters (α and β). These data is then utilized to train the machine learning models.

III. AUTOML METHODOLOGY

Fig. 3 presents the proposed AutoML architecture for smart characterization of inkjet printed components. The architecture leverages a suite of regression models and unfolds in several systematic stages, starting from data preprocessing to model training and evaluation, ensuring a comprehensive analysis of the inkjet printed components.

A. Data Processing

The data preparation is the first step of our AutoML architecture and comprises of several steps as follows:
• Data Collection: The training data includes frequency dependent $\alpha$ and $\beta$ by varying the material parameters ($\sigma_{ink}$, $\epsilon_{DS}$, $\epsilon_{FS}$ and $\tan\delta$).

• Data Description: The dataset consists of a substantial 47,200 data points for each parameter, with the frequency spanning a broad spectrum from 10 MHz to 20 GHz.

• Data Cleaning: We undertook a rigorous cleaning process to ensure elimination of discrepancies, duplicate values, thereby enhancing the overall quality of our data.

• Data Augmentation: We employed data augmentation techniques to synthesize new derivative data points, which in turn provides a more diverse and comprehensive dataset for training our models.

• Data Selection and Partitioning: We employed two distinct methods for partitioning the dataset. The first method involved a random selection process where we allocated 75% of the data for training, 5% for testing, and the remaining 20% for validation. The second method adjusted the distribution to utilize 90% of the data for training, with the remaining 10% split evenly between testing and validation, each receiving 5%.

B. Feature Engineering

This stage is crucial for enhancing the model’s predictive power and involves:

• Feature Creation: Generating new features from the existing dataset

• Transformations: Applying mathematical or statistical transformations to normalize or scale the data features.

• Feature Extraction: Deriving new features from the raw data that are more informative and non-redundant.

• Feature Selection: Selecting crucial features to reduce dimensionality and improve model performance.

• Missing Value Imputation: Handling missing data by employing techniques to input or generate substitute values.

C. Model Training

Subsequent to data preprocessing, the architecture embarks on the core phase of model training and evaluation. This phase employs a diverse array of 39 regression models, ensuring a comprehensive exploration of the solution space. The models include, but are not limited to, Decision Tree (DT), XgBoost (Fig. 4), LightGBM (Fig. 5), and ResNet. The
core of the AutoML architecture revolves around generating and refining machine learning models:

- Model Library: A repository of various machine learning algorithms, such as SVM, Decision Trees, XGBoost, LightGBM, ResNet, CNN, RNN, etc.,
- Hyperparameter Optimization: An iterative process of tuning the hyperparameters to find the most effective model configurations.
- Architecture Optimization: Fine-tuning the structure or architecture of models for optimal performance.
- Model Refinement: Iterative process for improving models based on the evaluation metrics and feedback.
- Monitoring: Continuous observation of model performance during the training and validation process.

The architecture emphasizes a feedback loop where the model evaluation enables further model generation, refinement, and feature engineering, ensuring a dynamic and adaptive AutoML system. In the Fig. 4 and 5 we see an exemplified workflow for a dataset undergoing processing, such as imputation of missing values, standardization, and subsequent training of machine learning models like XGBoost and LightGBM leading to predictions. This indicates a parallel processing and ensemble approach, possibly combining predictions from multiple models to enhance the final prediction performance. The presented AutoML architecture illustrates a sophisticated, iterative, and integrative approach to developing machine learning models.

D. Comparative Analysis and Model Selection

Upon the completion of training and evaluation, the architecture conducts a comparative analysis of the models. This analysis is pivotal for choosing the model that exhibits the most promising results in the context of smart additive manufacturing and includes:

- Low Fidelity: Rapid, approximate evaluations of model performance through training on subsets of data or reduced epochs.
- Weight Sharing: Utilizing shared weights across different models to reduce computational overhead.
- Early Stopping: Preventing overfitting by halting the training process when the model’s performance on the validation set plateaus or decreases.
- Transfer Learning: Applying knowledge from one domain to improve performance in another, often speeding up the training process.
- Statistical Tests: Using hypothesis testing to ascertain the statistical significance of the observed differences in model performance.
- Performance Metrics: Evaluating models using various metrics such as accuracy, recall, AUC-ROC for classification, and MSE, RMSE for regression.

The scatter plots shown in Fig. 6 delineate the predictive performance of various machine learning models, specifically DT, LightGBM, XGBoost, and ResNet, across four distinct parameters under study. The plots illustrate a comparative analysis between the actual and predicted values, where a perfect prediction would align with the blue line. For the parameter $\sigma_{ink}$, the XGBoost and LightGBM models demonstrate a tight clustering around the line of perfect prediction, indicating a robust predictive capability, while the DT model exhibits greater dispersion, signifying potential over-fitting or model variance. This trend is echoed in the $\epsilon_{FS}$ parameter, where all models show commendable predictive accuracy, although the DT model seems slightly more scattered compared to the ensemble methods. Interestingly, in the assessment of $\tan\delta$, the LightGBM and XGBoost models maintain their high fidelity to the actual values, with the DT model’s predictions diverging more noticeably at lower actual values, potentially indicating a lack of sensitivity to finer variations within the dataset.

The inclusion of the ResNet model in the $\sigma_{ink}$ parameter plot reveals its competitive performance alongside XGBoost, both adhering closely to the line of unity across a broad range of values, albeit with some deviation at higher magnitudes. The DT model, while generally following the trend, shows substantial variance at the higher end of the scale, suggesting that while it captures the overall pattern, its precision may be compromised under conditions of increased parameter values.

To further elaborate the comparative analysis, we report the root mean square error (RMSE) validation and holdout scores for regression models in Table I and Table II. The XGBoost model showcases commendable consistency, as indicated by identical $\sigma_{ink}$ and $\epsilon_{FS}$ scores of 0.2573, coupled with a rela-

![Fig. 6. Validation of training with holdout samples for retrieved material parameters using different models.](image)

<table>
<thead>
<tr>
<th>Model Name</th>
<th>$\epsilon_{DS}$</th>
<th>$\epsilon_{FS}$</th>
<th>$\sigma_{ink}$</th>
<th>$\tan\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGBoost</td>
<td>0.2573</td>
<td>0.306</td>
<td>0.2573</td>
<td>0.00379</td>
</tr>
<tr>
<td>LightGBM</td>
<td>0.00378</td>
<td>0.4171</td>
<td>0.3492</td>
<td>0.00379</td>
</tr>
<tr>
<td>DT</td>
<td>0.3492</td>
<td>0.4371</td>
<td>0.3492</td>
<td>0.00395</td>
</tr>
<tr>
<td>ResNet</td>
<td>7523400</td>
<td>0.4925</td>
<td>0.4358</td>
<td>0.00406</td>
</tr>
</tbody>
</table>
tively low RMSE validation score \( \tan\delta \) of 0.00379, suggesting a strong predictive accuracy with minimal overfitting. The LightGBM algorithm, while demonstrating a minuscule \( \sigma_{\text{ink}} \) coefficient of 0.00378, unexpectedly reports high \( \epsilon_{DS} \) and \( \epsilon_{FS} \) values at 0.4171, which, despite sharing an identical RMSE validation score with XGBoost, might point to an overestimation in prediction under certain data distributions or a potential overfitting that the RMSE score alone does not fully reveal.

The DT model presents a higher \( \sigma_{\text{ink}} \) coefficient and \( \epsilon_{FS} \) value at 0.3492 with the highest \( \epsilon_{DS} \) coefficient at 0.4371, aligning with a slightly increased RMSE validation score of 0.00395. This increment in RMSE suggests that the DT model captures a broader range of data patterns at the expense of prediction precision. The ResNet model, primarily recognized for its applications in deep learning, displays an excessively high \( \sigma_{\text{ink}} \) coefficient of 7523400, which is indicative of overfitting, especially when considered alongside the highest RMSE Validation Score of 0.00406 in the group. Its \( \epsilon_{DS} \) and \( \epsilon_{FS} \) scores of 0.4925 and 0.4358, respectively, though on the higher side, are overshadowed by the disproportionate \( \sigma_{\text{ink}} \) value, suggesting that the model’s complexity may not be justified for the given regression task [18].

In assessing the efficacy of various machine learning models for regression tasks shown in Table II, we observe a diverse range of outcomes as illustrated in the table of results. The XGBoost model exhibits a moderate level of precision indicating a relatively consistent and reliable prediction capability. In contrast, the RMSE holdout scores for LightGBM model suggest that it may be more sensitive to the specific dataset or require careful tuning of parameters to achieve optimal performance. The DT model shows a higher variation in coefficients. This could imply a model is more prone to overfitting or that it captures more complexity in the data. Lastly, the ResNet architecture, typically used for deep learning tasks, presents an anomaly in its \( \sigma_{\text{ink}} \) coefficient at 17025000, a clear indication of overfitting. This substantial \( \sigma_{\text{ink}} \) value might be a result of the model’s complexity and a large number of parameters, which could have led to capturing noise in the training data.

This comparative evaluation reveals that while XGBoost and LightGBM maintain a lower RMSE, indicating superior generalization, the DT shows a slightly reduced predictive precision, and the ResNet’s performance suggests an inclination toward overfitting, underscoring the importance of balancing model complexity with predictive accuracy in regression tasks. Overall, the varying performance metrics underscore the necessity for model-specific optimizations and highlight the trade-offs between model complexity and prediction accuracy.

### IV. Results and Analysis

Based on the comparative analysis of the RMSE validation scores of different machine learning models presented in section III, we chose XGBoost and LightGBM models for extracting the material parameters of the inkjet printed CPWs. Once the training was completed, we proceeded for extracting the material parameters from measured propagation constant. To this end, the attenuation, phase constant and frequencies were provided as input to the proposed AutoML model. Fig. 7 depicts the frequency dependent material parameters predicted from our proposed AutoML architecture. It can be seen that predicted \( \epsilon_{DS} \) vary between 1.7 and 2.2; whereas \( \epsilon_{FS} \) vary between 2.6 and 3.4. The predicted loss tangent of the flexible substrate is within 0.008 and 0.013. Finally, the conductivity of cu electroplated silver nanoparticle ink vary between 2.6 and 3.4. The extreme outliers for each prediction can be attributed to the worst case modeling error for the AutoML architecture and can be ignored. Nevertheless, we used the mean value from Fig. 7 to estimate the material parameters.

In order to verify the retrieved material parameters are correct, we measured \( \sigma_{\text{ink}}, \epsilon_{FS}, \epsilon_{DS}, \tan\delta \) using alternative approaches. A split postresonator method was adopted.

---

**Table II**

<table>
<thead>
<tr>
<th>Model Name</th>
<th>( \sigma_{\text{ink}} )</th>
<th>( \epsilon_{DS} )</th>
<th>( \epsilon_{FS} )</th>
<th>( \tan\delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGBoost</td>
<td>0.2582</td>
<td>0.3146</td>
<td>0.2582</td>
<td>0.00375</td>
</tr>
<tr>
<td>LightGBM</td>
<td>0.00376</td>
<td>0.4162</td>
<td>0.4162</td>
<td>0.00377</td>
</tr>
<tr>
<td>DT</td>
<td>0.3459</td>
<td>0.4517</td>
<td>0.3459</td>
<td>0.00389</td>
</tr>
<tr>
<td>ResNet</td>
<td>17025000</td>
<td>0.4898</td>
<td>0.4362</td>
<td>0.00401</td>
</tr>
</tbody>
</table>

![Fig. 7](image_url)

Fig. 7. Extracted material parameters from the proposed automated machine learning model: (a) \( \epsilon_{DS} \) (b) \( \epsilon_{FS} \) (c) \( \tan\delta \) (d) \( \sigma_{\text{ink}} \).

**Table III**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Measured</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon_{FS} )</td>
<td>3.2</td>
<td>3.05</td>
</tr>
<tr>
<td>( \epsilon_{DS} )</td>
<td>1.81</td>
<td>1.85</td>
</tr>
<tr>
<td>( \tan\delta )</td>
<td>0.01</td>
<td>0.0113</td>
</tr>
<tr>
<td>( \sigma_{\text{ink}}(10^7 \text{S/m}) )</td>
<td>2.973</td>
<td>3.107</td>
</tr>
</tbody>
</table>
Through extensive training and validation, we showed that the proposed automatic architecture eliminates the human expertise required to select appropriate algorithms. The extracted material parameters were found to be in good agreement with the measured values.

**CONCLUSION**

An automated machine learning based architecture has been proposed for smart characterization of inkjet printed components. Our proposed architecture evaluates numerous state-of-the-art machine learning models and automatically selects the best algorithm for extraction of each parameter. Through extensive training and validation, we showed that XGBoost and LightGBM methods offer better accuracy for the modeling and regression task under study. The proposed characterization methodology was able to accurately predict the ink conductivity and the dielectric properties of the flexible substrate and spacer used for measurement. The employment of machine learning based models enable full automation of the extraction procedure. Furthermore, the proposed AutoML architecture eliminates the human expertise required to select appropriate algorithms. The extracted material parameters were found to be in good agreement with the measured values.

**REFERENCES**


Fig. 8. Comparison of attenuation constant obtained from multiline TRL and EM simulation with material parameters predicted from our proposed machine learning model. The difference between measured and predicted attenuation constant is plotted on the right.