2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD) | 978-1-6654-9125-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIBD57115.2023.10206117

A Method of Predicting the Deposited Film Thickness in IC Fabrication Based on Automatic Machine Learning

Yumeng Shi
School of Micro-Nano Electronics
Zhejiang University
Zhejiang, China
22241023@zju.edu.cn

Yu Cai
School of Microelectronics
University of Science and Technology of China
Anhui, China
caiyu@mail.ustc.edu.cn

Yining Chen*
School of Micro-Nano Electronics
Zhejiang University
Zhejiang, China
yining.chen@zju.edu.cn

***Abstract*—The semiconductor IC (Integrated Circuits) industry involves a multitude of fabrication processes, and quality verification at critical stages is crucial for ensuring successful post-processing outcomes. However, conventional physical measurements can be time-consuming, and machine learning methods that predict thickness of deposited films require extensive effort in model selection and hyperparameter optimization. Automatic machine learning, a new intelligent learning approach that uses Bayesian optimization to automate feature extraction from input data, model selection and hyperparameter optimization, has the potential to minimize the time and cost required for training and model selection. This paper presents a novel approach for predicting film thickness of IC deposited films based on automatic machine learning. The study analyzes three key techniques including meta-learning, Bayesian optimization and model integration, and establishes a prediction model using the Auto-Sklearn system. The model's prediction performance is compared with three common machine learning algorithms, namely random forests, support vector machines and multilayer perceptron, and the results show that the automatic machine learning prediction model outperforms these algorithms with a coefficient of determination of 0.759 on the test set and a mean relative error of 0.262% for the prediction results. This paper demonstrates the efficacy of automatic machine learning for predicting film thickness and provides a promising direction for future work in the semiconductor industry.**

Keywords*—IC deposition film thickness, automatic machine learning, machine learning, prediction, feature processing

I. INTRODUCTION

In recent years, the IC industry has undergone significant advancements and developments. To evaluate the results of various critical processes such as etch rate, deposition rate, film thickness, chamber matching, chemical mechanical polishing, carrier distribution prediction, and uniformity, the IC sector requires metrology[1]. However, the traditional method of physical measurement is highly demanding in terms of time, equipment, and human resources, making it challenging to implement. Consequently, researchers have focused on developing models to predict these parameters as an alternative to physical measurement. In the field of IC manufacturing processes, there are two primary categories of prediction methods for process parameters: process mechanism-based and data-driven. However, some process mechanism-based models require the integration of intricate process mechanisms, which include numerous physical and chemical reactions, energy conservation equations, and so on. Due to the IC process's extensive structure and complicated and highly non-linear mechanisms, process modeling and control are extremely challenging. On the other hand, IC manufacturers have accumulated significant production process data with the development of smart instrumentation and computer technology, leading to the widespread adoption of data-driven forecasting methods in the IC industry. Traditional statistical data-driven methods are significantly influenced by the choice of model, whereas machine learning shows great promise for the prediction of various process parameters in IC manufacture due to its powerful data processing capabilities and the fact that no exact physical model or prior knowledge are required.

Over the past few years, researchers both domestically and internationally have conducted extensive studies on the use of machine learning to predict various process parameters in IC manufacturing[2]–[4]. However, there are still several shortcomings in the relevant research. Firstly, a complete feature selection principle has not been established, and feature selection is still limited to single-factor analysis of each influencing factor and the target variable, without considering the correlation between influencing factors. This results in a lack of objective feature selection and the retention of many useless features, which may weaken the generalization ability of the model. Secondly, conventional prediction models based on traditional machine learning methods require significant manual intervention and a substantial effort for algorithm selection and hyperparameter optimization of the model. As a result, the threshold of machine learning increases the difficulty and complexity of building prediction models in the IC field.

Automated machine learning techniques are capable of automatically extracting features and can be configured with different feature pre-processing methods, including feature selection, dimensionality reduction, generation, and coding, to produce models with optimal performance. Furthermore, these methods can be automated through Bayesian optimization, which automatically selects appropriate machine learning models and performs hyperparameter optimization. As a result, automated machine learning reduces the user's workload and

lowers the threshold for machine learning users. In this study, automatic machine learning is applied to build a prediction model for the deposited film thickness in IC fabrication The results show remarkable improvement comparing to ones predicted by conventional machine learning models[5], [6].

II. Research Methodology

A. General Machine Learning

To demonstrate the superiority of automatic machine learning methods, several common advanced machine learning methods have been selected for comparison, including random forests, support vector regression and multilayer perceptron. These three machine learning algorithms, which differ significantly in their algorithmic structure, have been tested over years to have qualify performance on most data sets. Modeling is performed using these algorithms, and the prediction performance of each algorithm is compared[7].

(1) Random Forest (RF): RF is a supervised learning algorithm, which is an integrated learning algorithm with decision trees as the base learner. The main parameters affecting the random forest algorithm include the number of base evaluators, the number of features considered when branching, the maximum depth of the tree, etc. By optimizing these hyperparameters, the prediction accuracy of the algorithm can be improved.

(2) Support Vector Regression (SVR): SVR algorithms are commonly used in the construction of regression prediction models. The main hyperparameters of the support vector regression model are the choice of kernel function, the insensitive loss coefficient, the penalty coefficient of the error term and so on, which are used as hyperparameters to be optimized.

(3) Multi-Layer Perceptron (MLP) is a convergent structured artificial neural network that consists of an input layer, an output layer and one or more hidden layers. The main parameters affecting MLP include the number of hidden layers and nodes, the choice of activation function, optimizer type, learning rate, batch size, etc. By optimizing these hyperparameters, the prediction accuracy of this algorithm can be improved.

B. Automatic Machine Learning

The automatic machine learning workflow is shown in Fig.1. The process consists of three main components, namely meta-learning, Bayesian optimization and model integration[8].

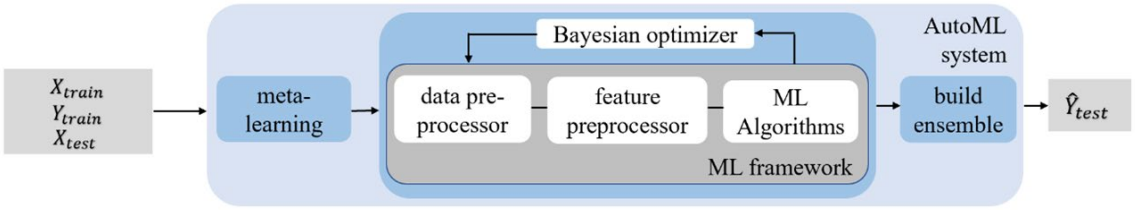

Fig. 1. Automatic machine learning workflow diagram.

(1) Meta-learning: Automated machine learning employing meta-learning as a precursor step to Bayesian optimization can significantly improve modeling efficiency. Meta-learning refers to the systematic observation of performance differences between existing learning tasks and then learning from existing experience and metadata, from which instances of machine learning frameworks that are likely to perform better on the present task are selected and used as initial parameters to perform the ensuing Bayesian optimization process[9], [10].

(2) Bayesian Optimization: Bayesian optimization, a method of using Bayes' theorem to guide the search to find the minimum or maximum value of an objective function, is the use of previously observed historical information at each iteration to perform the next optimization, resulting in a faster parameter search. The automatic machine learning approach used in this study automates the three parts of data preprocessing, feature preprocessing and algorithm engineering through Bayesian optimization.

(3) Model integration: Based on Bayesian optimization, automatic machine learning systems can obtain multiple models with good performance, and automatic machine learning frameworks can perform model integration on multiple trained machine learning models. Model integration can combine the learning capabilities of individual models to complement each other's strengths and improve the generalization capability of the final model and reduce the risk of overfitting. Integrated models typically outperform individual models. The automated machine learning system selects the model with the highest accuracy by weighting and summing the models to obtain an integrated model for parameter prediction.

(4) Automatic machine learning system: The Auto-Sklearn system was utilized in this study for automated machine learning. Auto-Sklearn is primarily based on the Sklearn machine learning library, making it easy for developers familiar with Sklearn to transition to Auto-Sklearn. In addition to better implementing the three techniques mentioned earlier, Auto-Sklearn also supports the inclusion of extended models and prediction processing methods. For instance, XGBoost algorithm support can be added to the machine learning models provided by Sklearn. Estimating model training times can be challenging based on personal experience, machine performance, feature size, data size, algorithm, and the number of iterations. Auto-Sklearn allows for the setting of single training times and overall training times,

enabling the tool to both limit training times and make the best use of time and computing power. Auto-Sklearn also supports a split training/test set approach and the use of cross-validation, reducing the amount of code and program complexity required for training the model.

III. Data Preparation and Analysis

A. Data Pre-processing

We obtained the CVD tool data and film thickness data from the production line of a real 12-inch IC foundry. The main parameters include parameters of the deposition process, the parameters of the pre-flow step, and the parameters of the preheating step, such as argon concentration, SIH_4 gas concentration, buffer pressure, wafer temperature and so on. We found that the data obtained directly from the machine had missing values, missing processes and multiple data for the same wafer, so we needed to pre-process the data first. And, to eliminate the effect of order-of-magnitude differences between features, the data set needs to be normalized with the following equation for the conversion function:

$$X^* = \frac{X-Min}{Max-Min} \tag{1}$$

where X^* denotes the normalized data, X denotes the sample data, Max denotes the maximum value of the sample data, and Min denotes the minimum value of the sample data.

B. Feature Filtering

After data pre-processing, we still have 68 features that are not filtered. In order to eliminate irrelevant features and improve the accuracy of our prediction, we propose a set of feature filtering principles to filter the features, i.e., the variance, outliers, and degree of linearity are considered together to remove variables. A small variance of a single feature indicates that the value of that feature does not vary much and is about a fixed value; a large proportion of outliers indicates that there are extreme values that may have a negative impact on the prediction of film thickness; and a degree of linearity close to 1 indicates that the two variables are strongly correlated and contain some overlap of information.

Box plots are plotted for each variable that affects film thickness. A box plot is a statistical chart used to show the dispersion of a set of data from which variables with too small a variance and too large a proportion of outliers can be filtered out by analyzing the box plot. The interval of each box line is from the 1st quartile Q_1 to the 3rd quartile Q_3 of the categorical data set, and the difference between Q_3 and Q_1 is defined as ΔQ. Outliers are defined as values greater than Q_3+1.5ΔQ or less than Q_1-1.5ΔQ.

Analysis of the box-line plots is shown in Fig. 2 and Fig. 3. By analyzing the box-line plots, features with variance less than 0.015 and features with a large proportion of outliers exceeding 0.2 were removed. In addition, the Pearson correlation coefficient was used to measure the linear correlation between the independent variables, and a heat map was drawn as shown in Fig. 4, from which features with a Pearson coefficient greater than 0.9 were screened out for deletion.

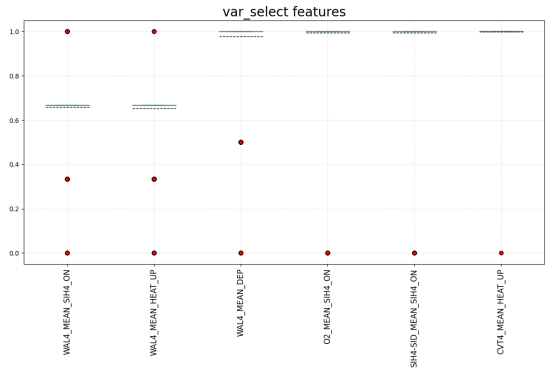

Fig. 2. Box line plots for variables with too small variance.

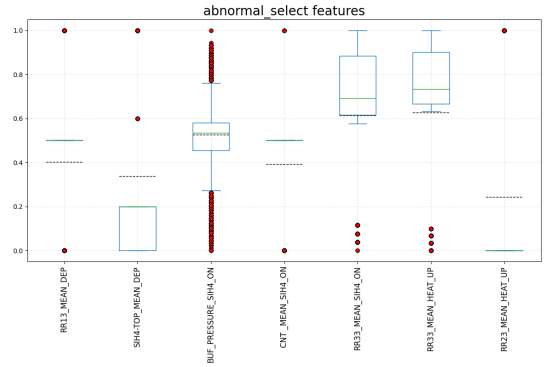

Fig. 3. Box line diagram for variables with disproportionately large outliers.

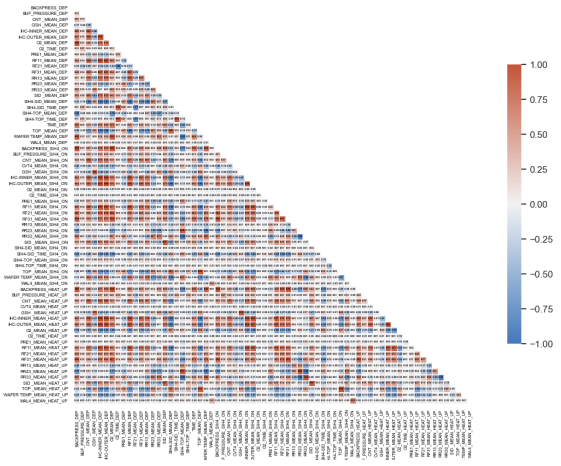

Fig. 4. Heat map of the correlation of each factor affecting film thickness.

Further, to compare feature screening results, the RF algorithm with n_estimators = 500 is used as a measure, and the coefficient of determination R^2 and mean square error MSE are used as evaluation criteria. The R^2 and MSE were compared for the five cases of not removing features, removing only features with too small a variance, removing only features with too large a proportion of outliers, removing features with too small a variance and too large a proportion of outliers, removing features with too small a variance, too large a proportion of outliers and features with a Pearson coefficient greater than 0.9 at the same time. The comparison results are shown in Fig. 5. Our proposed feature screening principle was shown to be

effective in achieving the best results by removing features with small variances, high proportions of outliers, and Pearson coefficients greater than 0.9.

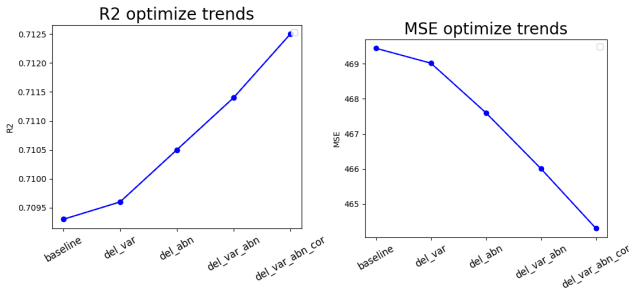

Fig. 5. Feature screening results.

IV. Machine Learning Prediction Model Building

A. Conventional Machine Learning Model Building

The Sklearn model package was used to implement the random forest algorithm, support vector regression and multilayer perceptron models. The Sklearn model package has various classification, regression and clustering algorithms, and integrates well with many other Python libraries. The dataset is randomly divided into a training set and a test set on a 75% and 25% scale. The training set is used for model training and hyperparameter optimization, while the test set is not involved in the training process and is only used to evaluate the predictive performance of the model. A 7-fold cross-validation approach is used to optimize the hyperparameters and select the optimal combination of hyperparameters to build the model. 7-fold cross-validation allows for efficient use of the data, robust evaluation of hyperparameter performance, reduction of model instability due to random partitioning of the dataset, and avoiding overfitting of the model.

Meanwhile, for the regression problem, we chose the coefficient of determination and the mean square error as indicators to assess the performance of the model, whose expressions are:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (3)$$

Where n is the total number of samples in the test data set, i is the sample number, $\hat{Y}_i$ is the predicted value of film thickness for the ith test sample, and Y_i is the true value of film thickness for the ith test sample. The prediction effectiveness of the various prediction models was evaluated by calculating their coefficients of determination and mean squared errors. The mean square error describes the difference between the predicted and true values of film thickness, and the smaller the value, the better the model is at predicting film thickness. The coefficient of determination is a statistical indicator of how reliably the regression model accounts for changes in the dependent variable, and the larger the value, the better the model predicts film thickness.

B. Automatic Machine Learning Model Building

Here we use the Auto-Sklearn library for automatic machine learning model building, where parameters such as training time, time allocated to each model evaluation, allowed models and feature pre-processing methods can be set separately. Prediction accuracy can be improved by extending the prediction time, but when the time is too long, it tends to increase the risk of overfitting, so we introduced a cross-validation method, using seven layers of cross-validation to reduce the risk of overfitting.

To evaluate the performance of automated machine learning in predicting IC process parameters, we limited the algorithm selection to three models: random forest, support vector regression, and multilayer perceptron, for comparison with traditional machine learning models. It is important to note that this restriction is only for the purpose of this study, and the automated machine learning approach is not limited to these three models.

V. Comparison of Prediction Results

A comparison of the prediction results of three conventional machine learning algorithms: random forest, support vector machine and multilayer perceptron with the automatic machine learning prediction model is shown in Table 1.

TABLE I. Prediction Results of Each Model

Algorithm name	Training Set			Test Sets		
	R^2	Mean Square Error	Relative Error /%	R^2	Mean Square Error	Relative Error /%
Automated Machine Learning	0.916	138.22	0.153	0.759	398.12	0.262
Random Forest	0.962	61.38	0.01	0.726	475.07	0.284
Support Vector Regression	0.829	277.87	0.048	0.723	479.57	0.291
Multi-Layer Perceptron	0.904	155.34	0.027	0.682	550.48	0.307

In terms of prediction accuracy, the automated machine learning prediction models show improved prediction results on the test set comparing to the traditional prediction models. The R^2 on the test set was 0.759, the mean square error was 398.12

and the mean relative error was 0.262%, which was better than other algorithms. Although the random forest model also showed good performance on the test set, its R^2 on the training set was too high, indicating that the model has a more serious overfitting phenomenon and weak generalization ability.

In terms of prediction efficiency, the use of traditional prediction models requires the empirical model selection and optimization of model parameters, a process that consumes too much time thus is inefficient in real production. Our automatic machine learning model, on the other hand, extracts feature from the input data without human help, select suitable machine learning models and automatically optimize the hyperparameters, which is faster to build and significantly reduces the time cost of investment.

Table 2 presents the parameters of each base predictor in the optimal automatic machine learning prediction model obtained with a training time of 15 minutes. The automatic machine learning model was ultimately chosen as the best predictor for IC deposition film thickness.

TABLE II. INFORMATION ON THE PARAMETERS OF EACH BASE PREDICTOR IN THE AUTOMATIC MACHINE LEARNING PREDICTION MODEL

Number	Regressor Type	Weighting ratio	Model Parameters	
			Parameter Name	*Parameter Value*
1	random_forest	0.44	n_estimators	512
			max_features	1.0
2	libsvm_svr	0.22	C'	194
			ε	0.001
			γ	0.201
3	mlp	0.2	hidden_layer_sizes	(100, 100, 100)
			learning_rate_init	0.0001
			max_iter	128
4	libsvm_svr	0.14	C'	1.427
			ε	0.100
			γ	0.057

VI. CONCLUSION

In this study, we propose an automatic machine learning approach to predict the deposition film thickness in the IC manufacturing process. The study introduces a set of feature screening principles that combine variance, outliers, and degree of linearity to simplify the model. By comparing the results of the automatic machine learning model with those of three traditional machine learning models, including random forest, support vector regression, and neural network, the study found that the automatic machine learning model outperformed the other algorithms with an R^2 of 0.759, mean squared error of 398.12, and average relative error of 0.262%. The Bayesian optimization algorithm used in automatic machine learning allows for automatic selection of appropriate machine learning models and hyperparameter optimization, leading to faster construction times and reducing the need for complex physical measurements. This approach can improve economic efficiency and provide guidance for the technical tuning of thin film deposition processes, reducing cycle times and costs.

ACKNOWLEDGMENT

- Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

- Competing interests

The authors have no relevant financial or non-financial interests to disclose.

- Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

- Authors' contributions

All authors contributed to the study conception and design. Yumeng Shi and Yu Cai completed the data analysis and code writing. Yumeng Shi designed and drafted the manuscript, and Yining Chen revised the paper. All authors read and approved the manuscript.

REFERENCES

[1] D.-H. Lee, J.-K. Yang, C.-H. Lee, and K.-J. Kim, "A data-driven approach to selection of critical process steps in the semiconductor manufacturing process considering missing and imbalanced data," *Journal of Manufacturing Systems*, vol. 52, pp. 146–156, Jul. 2019, doi: 10.1016/j.jmsy.2019.07.001.

[2] Z. Li, D. Wu, and T. Yu, "Prediction of Material Removal Rate for Chemical Mechanical Planarization Using Decision Tree-Based Ensemble Learning," *Journal of Manufacturing Science and Engineering*, vol. 141, no. 3, Jan. 2019, doi: 10.1115/1.4042051.

[3] A. Ferreira, A. Roussy, and L. Conde, "Virtual Metrology models for predicting physical measurement in semiconductor manufacturing," in *2009 IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, May 2009, pp. 149–154. doi: 10.1109/ASMC.2009.5155973.

[4] T. Zhou *et al.*, "Virtual Metrology of WAT Value With Machine Learning Based Method," in *2022 China Semiconductor Technology International*

Conference (CSTIC), Jun. 2022, pp. 1–2. doi: 10.1109/CSTIC55103.2022.9856736.

[5] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and Robust Automated Machine Learning," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2015. Accessed: Apr. 18, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/hash/11d0e6287202fced83f79975ec59a3a6-Abstract.html

[6] A. Kaul, S. Maheshwary, and V. Pudi, "AutoLearn — Automated Feature Generation and Selection," in *2017 IEEE International Conference on Data Mining (ICDM)*, Nov. 2017, pp. 217–226. doi: 10.1109/ICDM.2017.31.

[7] "Machine Learning: Algorithms, Real-World Applications and Research Directions | SpringerLink." https://link.springer.com/article/10.1007/s42979-021-00592-x (accessed Apr. 18, 2023).

[8] P. Ge, "Analysis on Approaches and Structures of Automated Machine Learning Frameworks," in *2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*, Jul. 2020, pp. 474–477. doi: 10.1109/CISCE50729.2020.00106.

[9] S. Dyrmishi, R. Elshawi, and S. Sakr, "A Decision Support Framework for AutoML Systems: A Meta-Learning Approach," in *2019 International Conference on Data Mining Workshops (ICDMW)*, Nov. 2019, pp. 97–106. doi: 10.1109/ICDMW.2019.00025.

[10] M. Gaikwad and A. Doke, "Survey on Meta Learning Algorithms for Few Shot Learning," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2022, pp. 1876–1879. doi: 10.1109/ICICCS53718.2022.9788260.