Supplementary Information for Non-destructive thickness characterisation of 3D multilayer semiconductor devices using optical spectral measurements and machine learning

Hyunsoo Kwak[1], Sungyoon Ryu[2], Suil Cho[2], Junmo Kim[3], Yusin Yang[2],
and Jungwon Kim*[,1]

[1] Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea
[2]Memory Metrology & Inspection Technology Team, Memory Manufacturing Technology Center, Samsung Electronics Co. Ltd., Gyeonggi-do 18448, Korea
[3] School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea

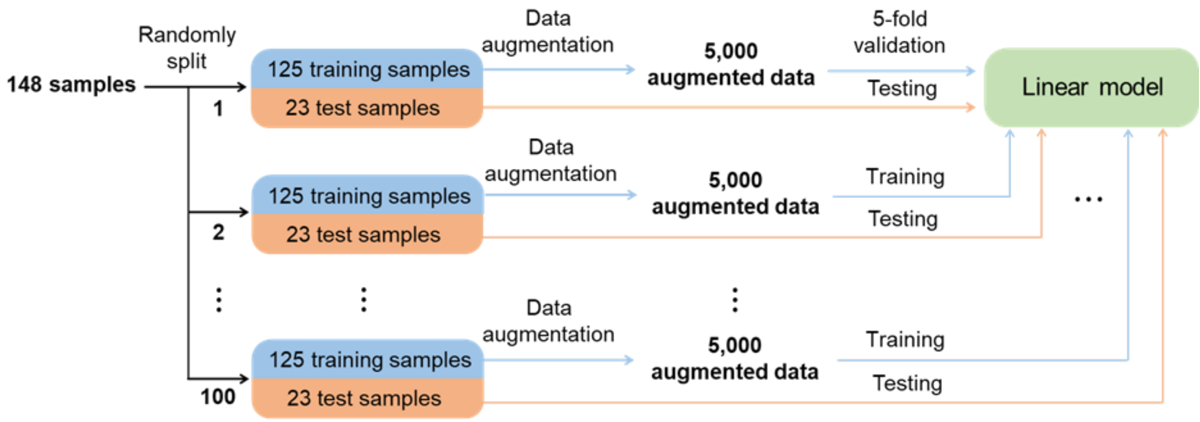

Figure S1 | Training and testing procedure for machine learning models for each-layer thickness prediction of normal condition samples. First, 148 samples are split into 125 training samples and 23 test samples, and then data augmentation is performed to generate 5,000 augmented data for model training. As a result of the evaluation of various machine learning models with five-fold cross-validation, the linear model shows the best performance. After 100 repetitions of random data splits, we could obtain the average thickness prediction results.

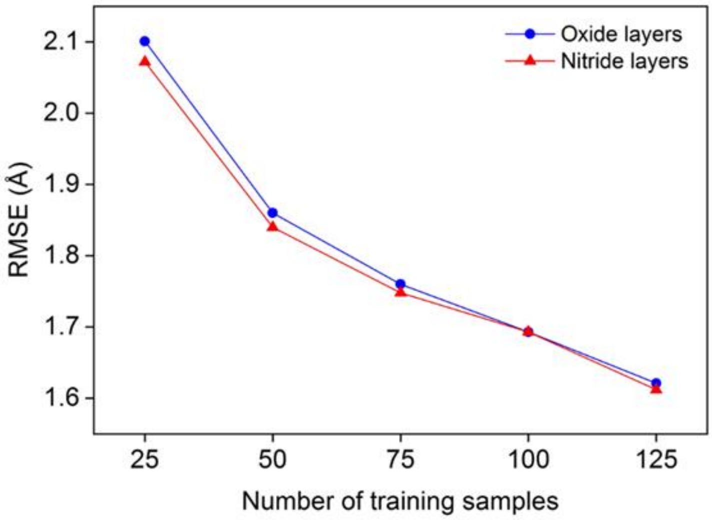

Figure S2 | Average prediction RMSE of test set. The average RMSE by modifying the number of training samples for 23 test samples over 100 repetitions of random data splits. Before training the linear model, each training sample is augmented by noise-injection method (40 augmented data per training sample).

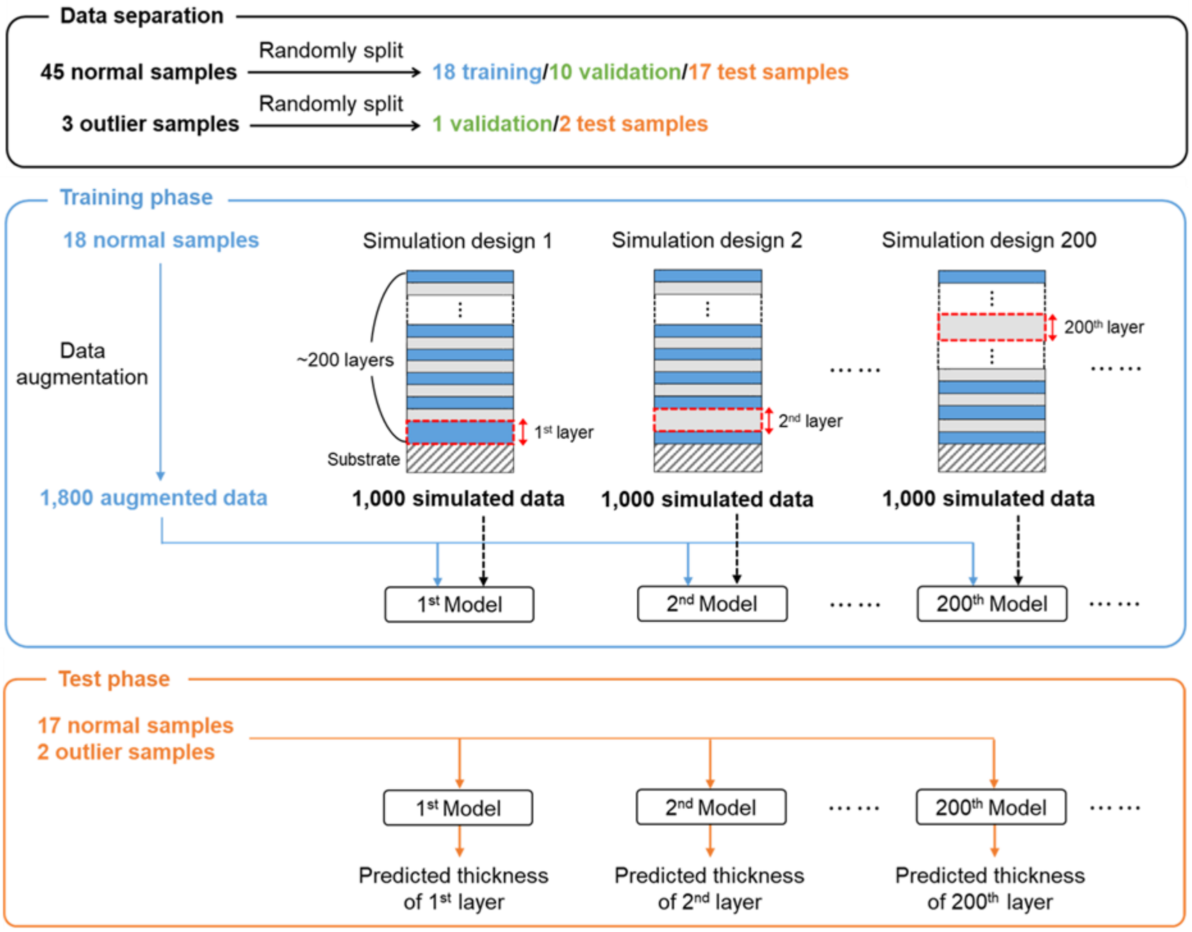

Figure S3 | Schematic of the outlier detection method. First, 45 normal samples are randomly split into 18 training, 10 validation, and 17 test samples. Three outlier samples are randomly split into 1 validation and 2 test samples. In training phase, each model is trained with 1,800 augmented data and 1,000 simulated data. The 1,000 simulated data are generated by simulation with a thickness profile that magnifies the thickness variation of each layer. In the test phase, 17 normal samples and 2 outlier samples are put into each model to predict each layer thickness. From the predicted thicknesses, we can determine whether the sample is in a normal condition, and for an outlier case, the thickness of the outlier layer can be obtained.

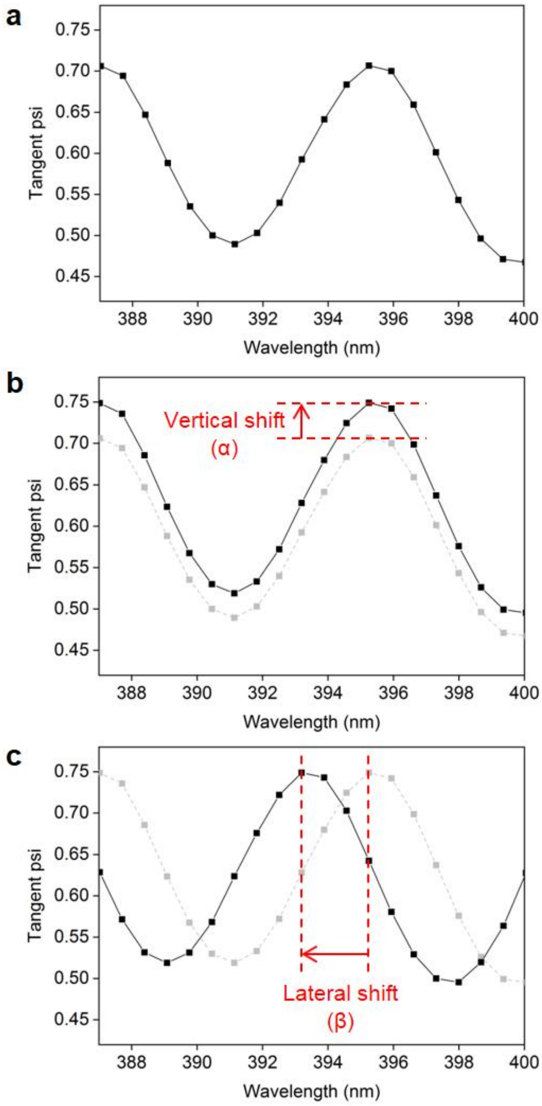

Figure S4 | Schematic of the noise injection method. a, Measured tangent psi of a normal sample. **b**, Vertical noise injection can be applied by adding α to the original spectral data. **c**, Lateral noise injection can be applied by interpolating the original spectral data into β-shifted spectral ranges.

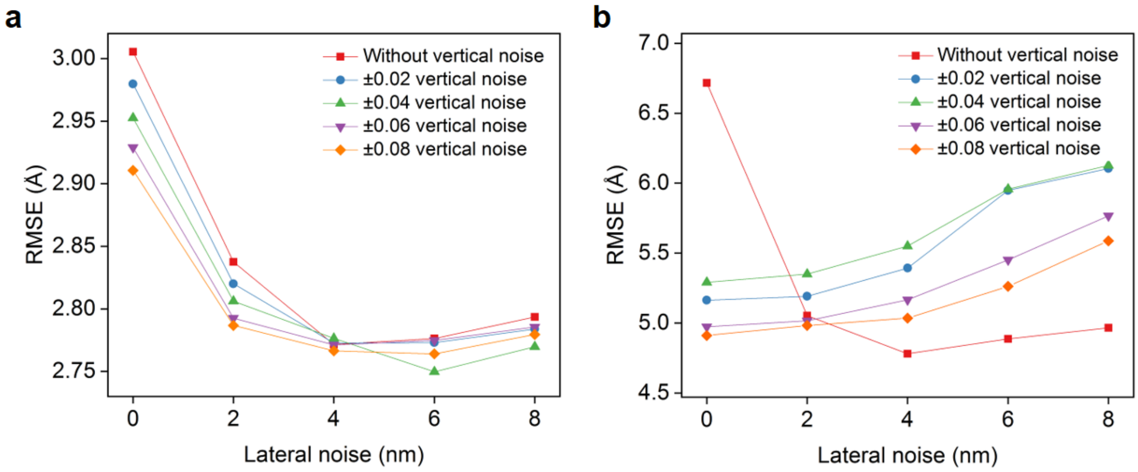

Figure S5 | Noise injection test in various conditions. Comparison of the RMSE of the validation set according to various amounts of lateral and vertical noise. The amount of lateral noise is modified from 0 (without lateral noise) to ± 8 nm (denoted by 8 in the horizontal axis). The amount of vertical noise is modified from 0 (without vertical noise) to ± 0.08. **a**, For multilayer metrology for normal conditions, the lowest RMSE is 2.75 Å when lateral noise with a uniform distribution of -6 nm to 6 nm and vertical noise with a uniform distribution of -0.04 to +0.04 is applied. **b**, For the outlier detection test, the lowest RMSE is 4.78 Å when lateral noise with a uniform distribution of -4 nm to 4 nm is applied without vertical noise injection.

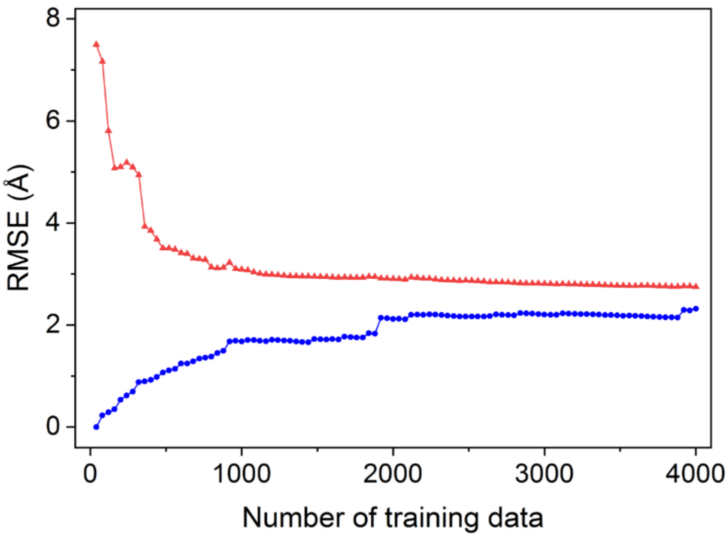

Figure S6 | Learning curve for the linear model. The linear model is evaluated with different numbers of training data. The number of training data is increased from 40 to 4,000 in 40 intervals (since each training sample is increased to 40 augmented data before being used for training), while the number of validation data is kept as 25. Blue circles denote the average RMSE of the training set over 5 folds, and red triangles denote the average RMSE of the validation set over 5 folds.

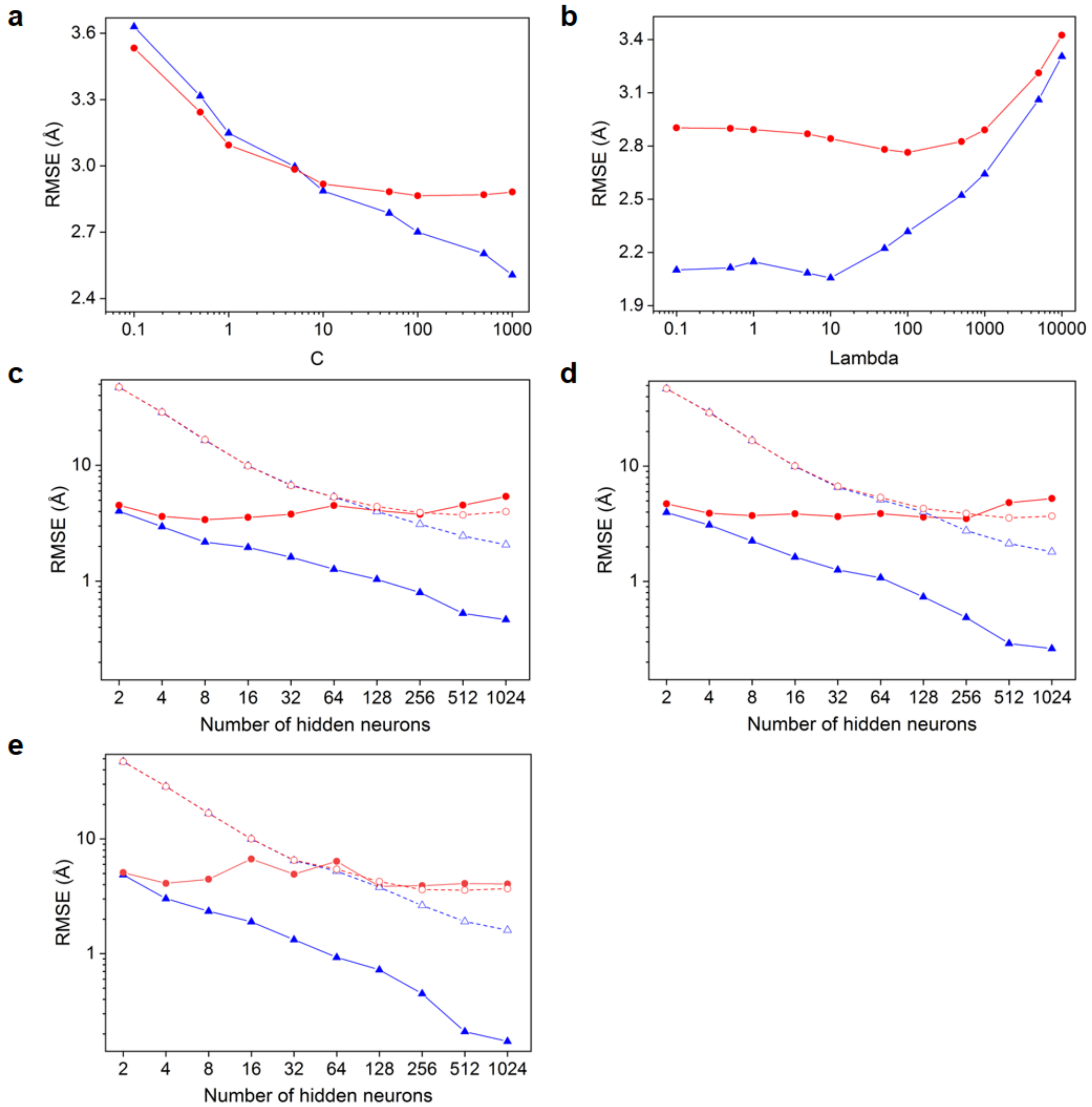

Figure S7 | Hyper-parameter tuning results for various models. To find the best model for multilayer thickness prediction for normal conditions, we compare the performance of each model by tuning various hyper-parameters. The RMSE is calculated by five-fold cross-validation. Blue triangles and red circles denote the RMSE of the training set and the validation set, respectively. **a**, RMSE according to regularization parameter C for the SVR model. **b**, RMSE according to the L2 regularization parameter, lambda, for the linear model. **c-e**, RMSE according to the number of hidden neurons for ANN models. We use a 2-layer NN (**c**), 3-layer DNN (**d**), and 4-layer DNN (**e**). Solid lines denote when dropout is not used at all; dashed lines denote implementation of dropout with a drop probability of 50% at the last hidden layer.

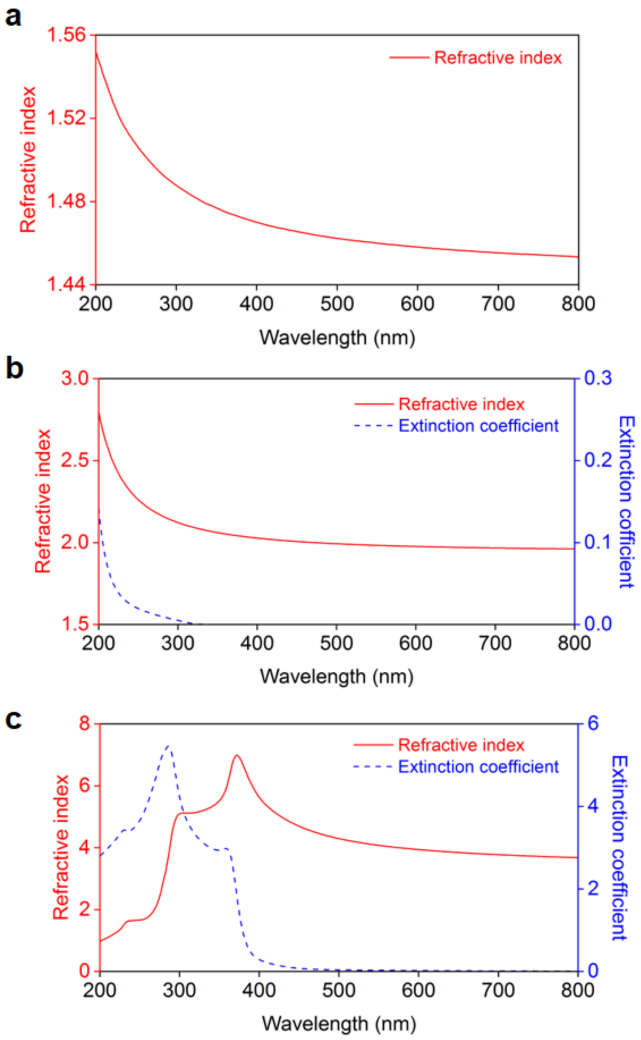

Figure S8 | Refractive index of materials. a, Refractive index of silicon oxide (low index material). The extinction coefficient for silicon oxide is zero for the range of 200 to 800 nm. **b**, Refractive index of silicon nitride (high index material). **c**, Refractive index of silicon substrate. All data are measured for a single layer material using an ellipsometer.

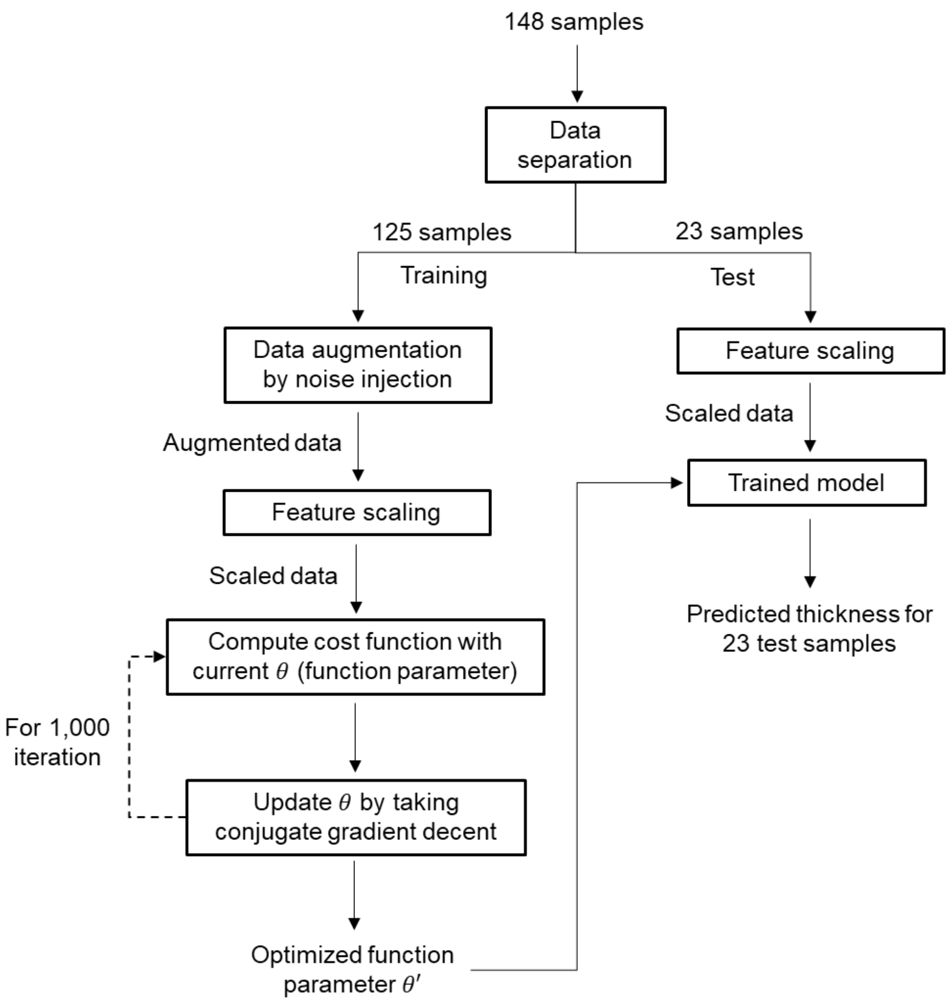

Figure S9 | Flow chart block diagram for the machine learning algorithm. Step-by-step algorithm operation process for training and test. Since we found that the linear model shows the best performance for thickness prediction, the parameter update in the training process is described as that of the linear model.

Table S1 | Thickness prediction results for four different structures.

	Structure 1	Structure 2	Structure 3	Structure 4
Total number of layers	~65	~110	~130	~200
Total thickness of design target	1.9 μm	3.6 μm	3.6 μm	5.5 μm
Spectroscopic data used for input [spectral range]	Psi [216 nm–905 nm], Delta [216 nm – 905 nm]	Reflectance [370 nm – 780 nm], Psi [200 nm – 980 nm], Delta [200 nm – 980 nm]	Psi [216 nm – 905 nm], Delta [216 nm – 905 nm]	Psi [216 nm – 905 nm], Delta [216 nm – 905 nm]
Number of training samples	65	60	65	125
Number of test samples	14	12	14	23
RMSE of oxide layers (standard deviation)	1.51 Å (0.2 Å)	1.59 Å (0.3 Å)	1.44 Å (0.2 Å)	1.62 Å (0.2 Å)
RMSE of nitride layers (standard deviation)	1.75 Å (0.4 Å)	1.68 Å (0.3 Å)	1.71 Å (0.4 Å)	1.61 Å (0.2 Å)

Table S2 | RMSE comparison of different combinations of psi and delta.

	RMSE of training set	RMSE of validation set
Psi only	2.48 Å	2.85 Å
Delta only	2.53 Å	2.85 Å
Psi and Delta	2.32 Å	2.75 Å

Table S3 | Five-fold cross-validation results for various models.

	RMSE of training set	RMSE of validation set
Support vector regression	2.70 Å	2.86 Å
Linear regression model	2.32 Å	2.75 Å
Neural network (2-layer)	2.18 Å	3.39 Å
Deep neural network (3-layer)	2.13 Å	3.55 Å
Deep neural network (4-layer)	1.91 Å	3.57 Å