

Machine learning approach to thickness prediction from *in situ* spectroscopic ellipsometry data for atomic layer deposition processes

FREE

Special Collection: [Atomic Layer Deposition \(ALD\)](#)

Ayush Arunachalam; S. Novia Berriel; Corbin Feit; ... et. al



Journal of Vacuum Science & Technology A 40, 012405 (2022)

<https://doi.org/10.1116/6.0001482>

CHORUS



CrossMark

Related Content

Calcium fluoride and gold reference by high sensitivity-low energy ion scattering

Surface Science Spectra (November 2019)

Crystalline anisotropy of cobalt ferrite from observations of ferromagnetic resonance

AIP Conference Proceedings (May 1976)

Growth mode evolution of hafnium oxide by atomic layer deposition

Journal of Vacuum Science & Technology A (November 2013)

■ Knowledge
■ Experience ■ Expertise

Click to view our product catalogue

Contact Hiden Analytical for further details:
www.HidenAnalytical.com
info@hiden.co.uk

Gas Analysis

- dynamic measurement of reaction gas streams
- catalysis and thermal analysis
- molecular beam studies
- dissolved species probes
- fermentation, environmental and ecological studies

Surface Science

- UHV TPD
- SIMS
- end point detection in ion beam etch
- elemental imaging - surface mapping

Plasma Diagnostics

- plasma source characterization
- etch and deposition process reaction kinetic studies
- analysis of neutral and radical species

Vacuum Analysis

- partial pressure measurement and control of process gases
- reactive sputter process control
- vacuum diagnostics
- vacuum coating process monitoring



Machine learning approach to thickness prediction from *in situ* spectroscopic ellipsometry data for atomic layer deposition processes

Cite as: J. Vac. Sci. Technol. A 40, 012405 (2022); doi: 10.1116/6.0001482

Submitted: 20 September 2021 · Accepted: 16 November 2021 ·

Published Online: 22 December 2021



Ayush Arunachalam,¹ S. Novia Berriel,² Corbin Feit,² Udit Kumar,^{2,3} Sudipta Seal,^{2,3,4,5,7}  Kanad Basu,^{1,a)} and Parag Banerjee^{2,4,6,8,a)} 

AFFILIATIONS

¹Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, Texas 75080

²Department of Materials Science and Engineering, University of Central Florida, Orlando, Florida 32816

³Advanced Materials Processing and Analysis Center, University of Central Florida, Orlando, Florida 32816

⁴Nano Science and Technology Center, University of Central Florida, Orlando, Florida 32816

⁵Biionix Faculty Cluster, College of Medicine, University of Central Florida, Orlando, Florida 32816

⁶Florida Solar Energy Center, University of Central Florida, Orlando, Florida 32816

⁷College of Medicine, University of Central Florida, Orlando, Florida 32827

⁸REACT Faculty Cluster, University of Central Florida, Orlando, Florida 32816

Note: This paper is part of the 2022 Special Topic Collection on Atomic Layer Deposition (ALD).

a) Authors to whom correspondence should be addressed: kanad.basu@utdallas.edu and parag.banerjee@ucf.edu

ABSTRACT

A machine learning approach is applied to estimate film thickness from *in situ* spectroscopic ellipsometry data. Using the atomic layer deposition of ZnO as a model process, the ellipsometry spectra obtained contains polarization data (Ψ , Δ) as a function of wavelength. Within this dataset, 95% is used for training the machine learning algorithm, and 5% is used for thickness prediction. Five algorithms—logistic regression, support vector machine, decision tree, random forest, and k-nearest neighbors—are tested. Out of these, the k-nearest neighbor performs the best with an average thickness prediction accuracy of 88.7% to within ± 1.5 nm. The prediction accuracy is found to be a function of ZnO thickness and degrades as the thickness increases. The average prediction accuracy to within ± 1.5 nm remains remarkably robust even after 90% of the (Ψ , Δ) are randomly eliminated. Finally, by considering (Ψ , Δ) in a limited spectral range (271–741 nm), prediction accuracies approaching that obtained from the analysis of full spectra (271–1688 nm) can be realized. These results highlight the ability of machine learning algorithms, specifically the k-nearest neighbor, to successfully train and predict thickness from spectroscopic ellipsometry data.

Published under an exclusive license by the AVS. <https://doi.org/10.1116/6.0001482>

I. INTRODUCTION

Atomic layer deposition (ALD) exploits a sequential set of two or more surface reactions repeated cyclically to achieve thin film growth with subnanometer resolution.¹ While new chemistries^{2,3} and applications^{4–6} are rapidly growing, the understanding of atomistic processes that occur on the substrate surface at length scales that are of engineering importance are crucial for advances in ALD process development and its accelerated adoption in

diverse industries. The use of *in situ* techniques plays a crucial role in achieving this objective. Techniques such as quartz crystal microbalance,^{7–9} Fourier transform infrared spectroscopy (FTIR),^{10–13} x-ray photoelectron spectroscopy,^{14,15} scanning tunneling microscopy,^{16,17} electrical resistivity measurements,^{18,19} x-ray synchrotron and fluorescence,²⁰ quadrupole mass spectroscopy,^{9,21} microcalorimetry,²² and spectroscopic ellipsometry (SE)^{23,24} are employed to reveal mechanistic aspects of growth in ALD.

Among the above techniques, SE is a powerful approach for *in situ* monitoring of ALD processes and thin film processes,²⁵ in general. SE uses the change in polarization of an elliptically polarized incident light to obtain information on the refractive index and thickness of the film being deposited. Using standard optical models of the substrate and film, the thickness of the film can be obtained in real time. This information helps us to determine the growth rate (i.e., thickness/cycle) of the ALD process. The growth rate is a key parameter to determining the three defining characteristics of an ALD process, i.e., (i) self-saturation, (ii) temperature stability window, and (iii) linear growth rate. Besides, *in situ* SE also provides sensitivity to minute and dynamic variations to thickness during each ALD half-cycle, signifying the gain in physical thickness and loss of ligands as gaseous by-products.²³ Thus, the ability to obtain the growth rate in real time makes SE an insightful *in situ* technique for studying and rapidly optimizing ALD processes.

The success of SE lies in the ability to use a physics-based optical model containing the refractive index (n) and extinction coefficient (k) for the film/substrate combination of interest. These are supplied by manufacturers of the SE equipment. This modeling process is shown in Fig. 1(A), and the counter i updates the values of the model parameters defined in step 2 such that the mean squared error is reduced through each iteration. The complexity of the software, including fitting algorithms and associated modeling tools, can provide a significant boost in the ability of a researcher to understand new ALD processes or determine when ALD processes deviate from expected behavior. While the importance of SE software-based analyses cannot be understated, it is to be noted

that SE, at its most basic level, is a measurement technique where the raw data is fitted to established models to obtain physical parameters of interest.

An alternate approach in the use of SE data can be the following. If optical models (or dispersion relationships of n and k of films and substrates) for the fit are known *a priori*, then it is possible that using machine learning (ML) algorithms, the SE data can be “trained” to directly extract thickness *without* the need for optical modeling. This idea is shown schematically in Fig. 1(B). Here, we stress that we do not propose to negate the use of the optical models, but merely use the model to train SE data to predict thicknesses in the most efficient way, rapidly and with consistent accuracy. Such a situation might be encountered in metrology for advanced manufacturing where film compositions are known and throughput has to be maximized. The optical model then simply serves as the backbone for training the SE dataset. In a larger context, we see our work as a first step toward a database of ML models that can help predict n and k dispersion relationships, when complex compositions with no prior experimental data are synthesized [Fig. 1(C)].

ML-based algorithms are being increasingly used to analyze characterization data such that a material’s “fingerprint,” i.e., a set of descriptors, can be mapped to its properties.²⁶ For example, ML-based approach has been used to analyze Raman,²⁷ FTIR,²⁸ and x-ray diffraction²⁹ data for deciphering complex spectra from multicomponent materials. Accordingly, there has been one investigation recently³⁰ looking at SE data using an ML scheme. Here, the authors supplement ellipsometry data with transmission and reflection spectra to arrive at a film’s physical parameters of interest

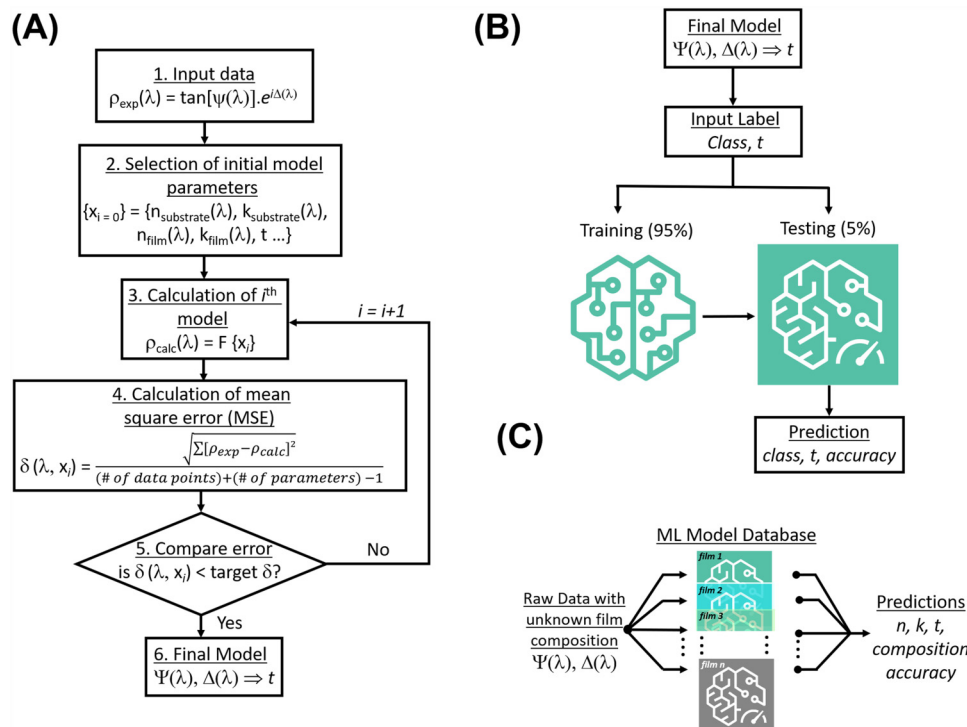


FIG. 1. (A) Generalized algorithm for analyzing SE data [Adapted with permission from E. A. Irene, *In Situ Real Time Characterization of Thin Films*, edited by Orlando Auciello, Alan R. Krauss, and Eugene Irene (Wiley, New York, 2001). Copyright 2001, Clearance Center, Inc.], where ρ is the complex reflection coefficient, λ is the wavelength, Ψ is amplitude ratio, and Δ is phase shift. (B) Our approach adopts the SE model from (a), classifies the dataset, and trains 95% of the data to correlate Ψ and Δ to thickness, t . The effectiveness of the ML model is ascertained by testing it on the remaining 5% of the raw data. (C) The model is now ready to accept any Ψ and Δ of the film and predict its thickness t with a certain accuracy.

namely, n , k , and thickness. The approach takes advantage of the already available, large database of optical constants for 200 materials and creates an iterative scheme for error minimization between the experimental data and the one predicted by the ML model. By necessity of the number of unknowns, the above approach relies on two measurement modalities—SE and reflectometry (and related transmission measurement) to fully define the problem. Obviously, the transmission measurements pose a challenge for opaque substrates. Furthermore, it is not known how such an approach might work for nonstandard films, films with interfaces, and mixed compositions.

In this paper, we use ML-based algorithms to predict thickness solely from *in situ* SE data of ALD ZnO, which are *a priori* modeled for its thickness. The optical modeling of the SE data provides us the ZnO thickness as a function of deposition time and serves as the training set for the ML algorithm. While the present analysis is focused on single film chemistry (i.e., ZnO) with a quasi-continuous range of thickness, a similar approach can be adopted to any film of interest for which there exists a SE model that the ML algorithm can be trained upon. Here, we note that the quasi-continuous nature of the ALD thickness data is important since it allows the ML algorithm to be trained across a wide range of finely spaced thickness values without having to resort to individual, discrete experiments. In this context, the use of ALD-based *in situ* data is particularly advantageous.

This paper is divided into Secs. II–IV. The experimental section in this paper describes the hardware setup used for *in situ* acquisition of SE data in ALD followed by a brief but pertinent introduction to SE. Next, the ML approach is described in detail. The results section in this paper highlights the effectiveness of using various ML algorithms in predicting the ZnO film thickness, strategies for improving prediction accuracy to within ± 0.5 nm and finally, the effect of downsampling the SE data, either randomly or via screening fixed regions in the SE spectrum.

We show that ML-based algorithms, when optimally trained, can effectively predict the thickness of ALD ZnO films to within ± 1.5 nm with over 87% accuracy. Furthermore, it is not necessary to have the entire spectrum of the SE data available. The predictive capability of the ML algorithm remains remarkably robust even when only 10% of the data are randomly made available or when SE data are present in a narrow spectral range. This result showcases the redundant nature of SE data when an ML algorithm is adequately trained. Taken together, these results highlight the utility of an ML-based approach to removing modeling complexity, reducing computational overhead costs, and ultimately, easing hardware requirements for *in situ* SE, specifically for ALD, and for thin film processing in general.

II. EXPERIMENTAL DETAILS

A. Atomic layer deposition

ZnO was deposited using a FIJI Gen2 ALD system from Veeco® using alternate pulses of diethyl zinc (DEZ, Sigma Aldrich, CAS:557-20-0 ≥ 52 wt. % in hexane) and de-ionized water (H_2O). The precursor pulses consisted of 0.06 s of DEZ and H_2O , separated by 6 s of inert gas (Argon) purge. The temperature of deposition was 150 °C. A Si substrate with native oxide, ultrasonically cleaned in a

mixture of H_2O and isopropyl alcohol and compressed-air dried, was used as a substrate. A total of 236 cycles were applied to grow the ZnO film that took approximately 51 min to complete. The total film thickness was 44.08 nm, and the continuous film growth rate was recorded as an average value of 0.187 nm/cycle using *in situ* ellipsometry, details of which are provided below.

B. *In situ* spectroscopic ellipsometry

To track the thickness of the sample, an M-2000 spectroscopic ellipsometer (SE) from J. A. Woollam® was used. The light source and detector were attached to the FIJI Gen2 ALD system on specialized mounts on 2 $\frac{3}{4}$ -in. conflat flanges with quartz windows, internally purged with 50 SCCM of argon gas. The angle of light incidence and reflection were 69.5°. The setup is shown in Fig. 2. The wavelength range for data acquisition was 271–1688 nm and consisted of 661 data points per scan. The acquisition time for each spectrum was ~ 3 s. This translated to a total of 1113 scans for an ALD process time that lasted 51.17 min.

The CompleteEase software by J.A. Woollam® was used to create an optical model for ZnO on Si. The native oxide on Si was estimated by the model to be 1.66 nm. A general oscillator model was used to model the ZnO film where the complex dielectric function ($\epsilon_1 + i\epsilon_2$) was modeled as a combination of a UV pole (ϵ_1) and a single gaussian peak (ϵ_1).³¹ While the background theory of SE is covered in textbooks^{25,32} and review articles^{23,24,33} that detail the use of SE as an *in situ* technique, here we briefly explain the nature of the data obtained via SE. For each scan, the polarization quantities ψ — Ψ and δ — Δ constitute the raw data obtained from the SE. The complex reflection coefficient, ρ , is then given as

$$\rho = \frac{r_p}{r_s} = \tan(\Psi)e^{i\Delta}, \quad (1)$$

where r_p and r_s are the parallel and perpendicular Fresnel coefficients, respectively, and are the ratios of the electric field with

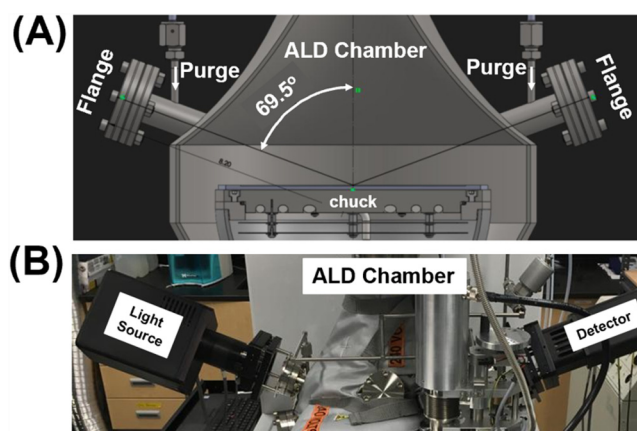


FIG. 2. (A) Engineering cut-out of the schematic for *in situ* ellipsometry setup on the FIJI Gen2 system (image courtesy Veeco®). (B) The M2000 J. A. Woollam® source and detector arms attached to the FIJI Gen2 chamber.

respect to the plane of incidence; Ψ is the amplitude ratio, and Δ is the phase shift. For multiple films, the Fresnel coefficient of the film stack, R_p , is given as (shown here for the parallel p component)

$$R_p = \frac{r_{p,\text{air-film}} + r_{p,\text{film-substrate}} e^{-i2\beta}}{1 + r_{p,\text{air-film}} r_{p,\text{film-substrate}} e^{-i2\beta}}, \quad (2)$$

where $\beta = 2\pi(\frac{t}{\lambda})N_{\text{film}} \cos(\varphi_{\text{film-substrate}})$, t is the thickness of the film, λ is the wavelength of the light, N_{film} is the complex refractive index (i.e., $N_{\text{film}} = n + ik$) of the film, and $\varphi_{\text{film-substrate}}$ is the angle of incidence of the light on the film-substrate interface. Using this approach, the optical constants (n and k as a function of λ) and t can be determined.³²

For ALD ZnO, Ψ and Δ as a function of wavelength are shown in Figs. 3(A) and 3(B), respectively, sampled at discrete time intervals during the ALD process. The time interval between each of the consecutive curves is approximately 8 min. It can be observed that there is a gradual change to the Ψ (increases) and Δ (decreases) as the ZnO thickness (i.e., time) increases. The frequency of data collection, however, is much higher (i.e., ~20 times/min). Therefore, the data can be plotted as a quasi-continuous contour plot of Ψ and Δ as shown in Figs. 3(C) and 3(D), respectively, where there are 1113 scans each of Ψ and Δ in the 51.17 min that it took for the process to complete. The dataset of Ψ and Δ are shown as a function of wavelength (x axis) and as a function of deposition time (y axis). Using the CompleteEase software from J. A. Woollam*, the Ψ and Δ measurement data can be translated into the thickness of ZnO. This is shown in Fig. 3(E) where the inset schematic shows the model film/substrate “stack” used. It is observed that the ZnO film shows an incubation period of ~15 cycles for growth³⁴ even on Si with native oxide (SiO_x).

Section II B describes the physics-based approach that is conventionally adopted in estimating thickness from SE measurements. However, to adopt an ML strategy for predicting thickness from Ψ and Δ , the data must be first curated and then trained using various ML-based algorithms. It should be noted that the robustness of the machine learning model is closely intertwined with that of the physics-based model. Thus, as in the physics-based model, a new model must be generated for each film stack composition. Once the algorithm is trained, however, any stacks of the same arrangement can utilize the same trained ML model for thickness prediction.

These details are provided next.

C. Curated dataset

The measured ZnO film thickness data are quasi-continuous in nature till a final thickness of 44 nm. Therefore, we first discretize these data by considering the integer values of thicknesses as shown in Fig. 4, which is a repeat of Fig. 3(E) but demarcated with the classification strategy. Classification is achieved by defining a step-size of 0.2 nm and selecting films that range between x and $x + 1$ nm. For example, films of thickness values 1, 1.2, 1.4, 1.6, and 1.8 nm are grouped together and considered as 1 nm. Subsequently, these integer film thicknesses are binned into 15 classes by employing a user-defined threshold of 3 nm. Hence, each class comprises three subclasses (for three thickness values). For example, class 0

consists of thickness values 1, 2, and 3 nm, while class 1 encompasses 4, 5, and 6 nm films. Each class consists of three thickness values, where each thickness corresponds to multiple batches of 661 Ψ and Δ values ranging from 271 to 1688 nm. Next, we leverage this multiclass dataset to train various ML algorithms toward a multiclassification task. Subsequently, a two-level classification strategy is adopted, where the first level, called *level-1 classification*, is associated with predicting which among the 15 classes a pair of Ψ and Δ values belong to, thereby having a granularity of 3 nm. On the other hand, the second level of classification, called *level-2 classification*, is associated with increased granularity of 1 nm and the thickness value which best describes a set of Ψ and Δ values.

D. Evaluation strategy

The ML approach is evaluated on five algorithms given as (i) k-nearest neighbors (kNN), (ii) random forest (RF), (iii) decision tree (DT), (iv) support vector machines (SVM), and (v) logistic regression (LR) using the curated multiclass data. We use the above selected algorithms because of their proficiency and efficiency in classification tasks involving multiple classes.³⁵ Furthermore, these algorithms are robust, immune to overfitting during training, and require minimal to no prior knowledge about the input data distribution. As opposed to binary classification, which involves two classes, multiclass classification is associated with classifying data points into one among a range of classes, which aligns with our objective of classifying Ψ and Δ into one among 15 classes. Modeling such a multiclass classification task predominantly involves prediction of Multinoulli probability distribution for all data points, which, in our context, is equivalent to predicting the probability of a pair of Ψ and Δ points belonging to each class.

Since the Ψ , Δ , and thickness correlations have been established in the previously shown data in Fig. 3, it is convenient to train the models. We use 95% of the data for training and use the remaining 5% to assess, where both the 95% and 5% of the data are chosen randomly. A two-level evaluation is performed, where the efficacies of classifying datapoints into different classes and subclasses are examined. The class- and subclass-level scores are labeled as *level-1 accuracy* and *level-2 accuracy*, respectively. *Level-1 accuracy* describes the efficacy of classifying datapoints into one of the 15 classes and has a granularity of 3 nm since each class comprises three subclasses (corresponding to three thicknesses). *Level-2 accuracy*, on the other hand, is associated with an increased granularity of 1 nm and represents the effectiveness of predicting film thickness, down to 1 nm. To better comprehend model performance, we analyze classification accuracy scores at different thicknesses.

III. RESULTS AND DISCUSSION

A. Prediction accuracy is impacted by choice of ML algorithm

The first case study compares the performance of the aforementioned ML algorithms as summarized in Fig. 5. The percent accuracy (Y axis) illustrates the probability of each model to classify a datapoint into one of the 15 classes of the curated dataset to within ± 1.5 nm accuracy, i.e., *level-1 classification*. kNN furnishes the best accuracy score of 87.5%. Here, each datapoint of the test

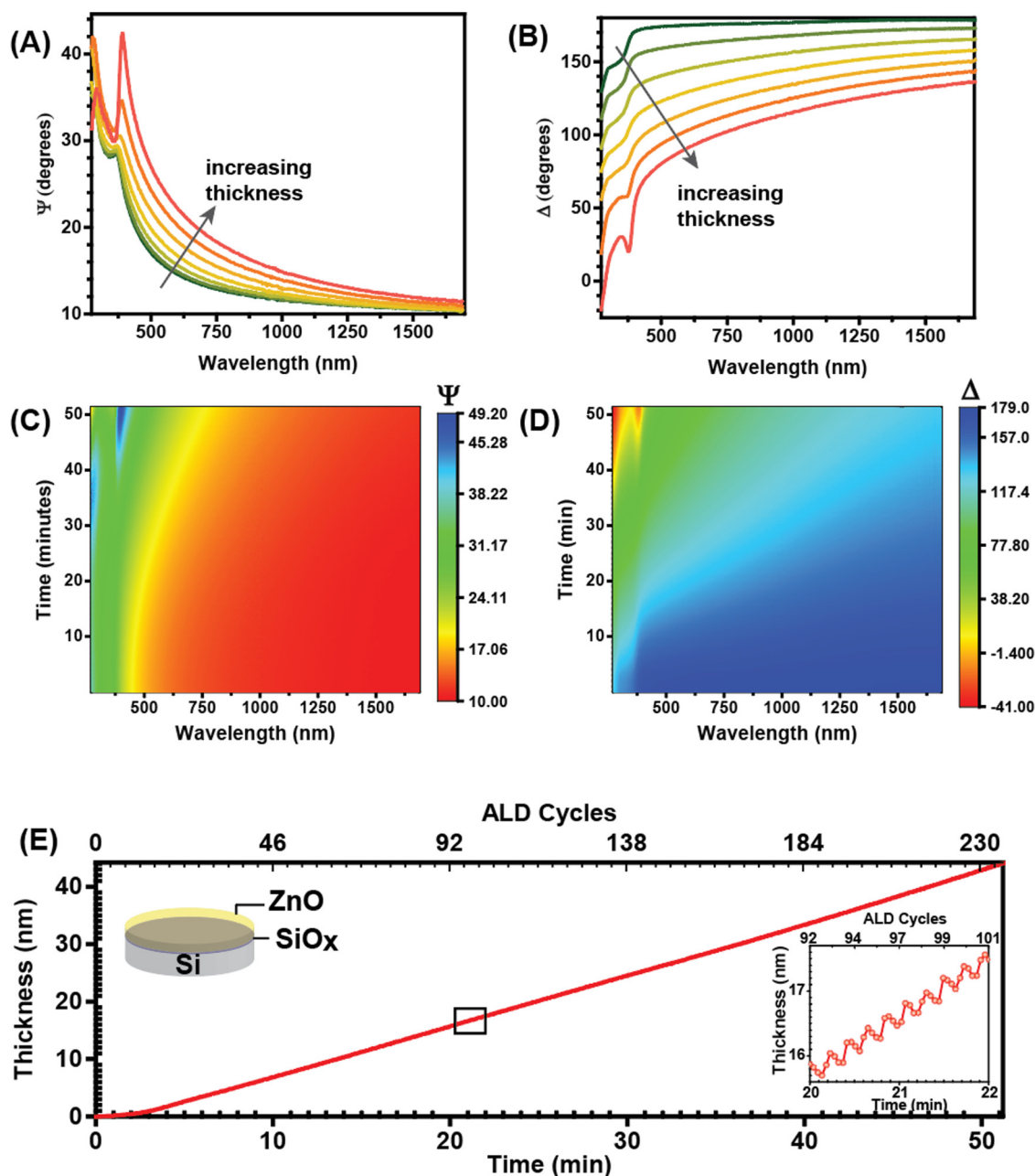


FIG. 3. (A) Ψ as a function of wavelength for various times from 0 to 51 min with 8-min intervals. (B) Δ as a function of wavelength for various times from 0 to 51 min with 8-min intervals. (C) Contour plot of the Ψ shows continuous variation with time (y axis). (D) Contour plot of the Δ shows continuous variation with time (y axis). (E) Thickness vs time (lower x axis) and cycle number (upper x axis), derived from the Ψ and Δ ellipsometry data. The schematic of the optical model used is shown in the top-left inset, whereas the bottom-right inset shows a zoomed image of the thickness behavior as a function of time/ALD cycles.

set is classified with a confidence of 87.5% using five neighbors.^{36,37} SVM, DT, and RF report accuracy results of 83.8%, 84.5%, and 86.8%, respectively. SVM uses a radial basis function of degree three as its kernel along with a regularization coefficient of value 1.

RF is an ensemble of decision trees, and like DT, requires a minimum of two samples to split an internal node. The maximum depth of the tree is not specified in either algorithm, implying that the tree can expand until all datapoints are accounted for. In this

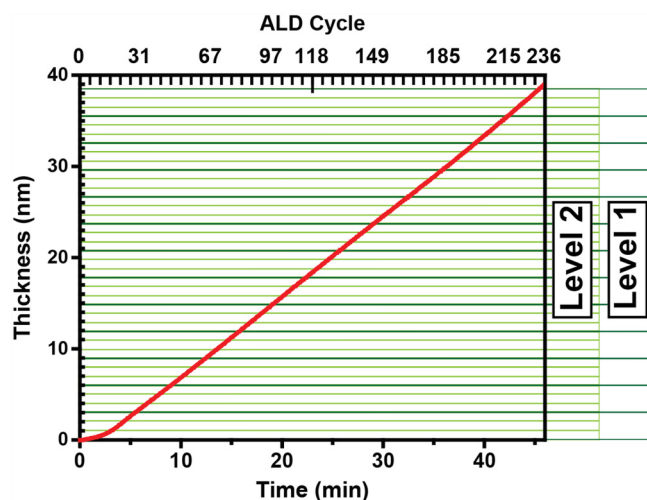


FIG. 4. Thickness vs time (lower x axis)/ALD cycle numbers (upper x axis) data are shown on the right to highlight level-1 and level-2 classifications. Level-1 class consists of thicknesses in sets of 3 nm each, whereas level-2 sub-class divides level-1 into three subsets of 1 nm each.

experiment, we consider 40 trees in RF. LR, on the other hand, uses a regularization coefficient of value 1 and limited-memory Broyden–Fletcher–Goldfarb–Shanno (*lbfgs*) optimization technique^{36,37} to calculate the parameter weights that minimize the cost function.

LR performs the worst among all the models, producing a score of 40.7%. This can be attributed to LR being inept at solving nonlinear problems owing to its linear decision surface, the requirement of having an exhaustive dataset, and its sensitivity of outliers in the dataset. Owing to its low accuracy score, LR will not be considered for future experiments. Additionally, SVM too is excluded from subsequent case studies for two reasons. The first reason is attributed to its high training time, while the second to its performance.

B. Prediction accuracy is impacted across classification levels-1 and -2 and film thickness

The second case study evaluates the performances of the three best algorithms, DT, RF, and kNN, with increasing granularity in the prediction of film thickness. These results assume importance from a process perspective since it is important to understand the uncertainty in accuracy (i.e., delta thickness) that is inherently present in an ML algorithm and the uncertainty that is propagated as a function of the true physical thickness of the film. Accordingly, accuracy scores of each model are as illustrated in Fig. 6(A) at both *level-1* and *level-2*. We have updated the hyperparameters of the models to bolster classification performance of each model. Increasing the number of neighbors from 5 to 21 improves the performance of kNN from 87.5% (as shown in Fig. 5) to 88.7% resulting in an enhancement of 1.2%. The number of trees in RF is increased from 40 to 100, which furnishes an

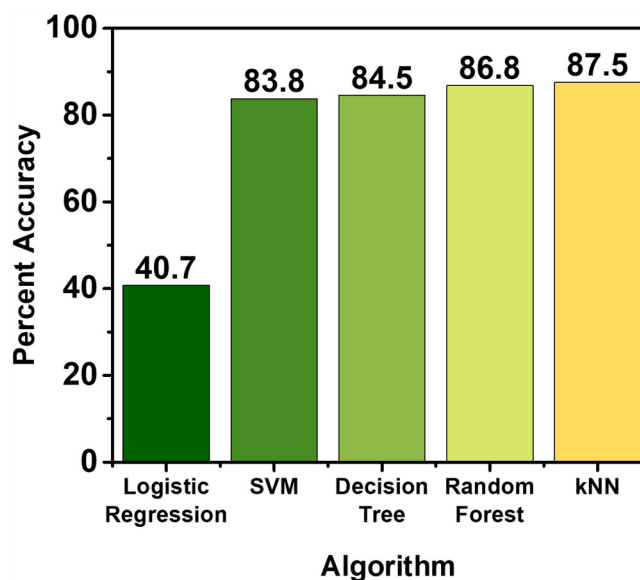


FIG. 5. Accuracy results of various ML algorithms for collated *in situ* ZnO ellipsometry data. This figure illustrates the confidence of each algorithm to classify a particular *in situ* data point into one of the 15 main classes of the dataset. Logistic regression and kNN furnish the lowest and highest scores of 40.7% and 87.5%, respectively. SVM, decision tree, and random forest yield scores of 83.8%, 84.5%, and 86.8%, respectively.

accuracy of 87.3% vs 86.8%. DT retains the same set of hyperparameters as before, which were optimized to produce best performance, and, thus, produces the same accuracy score of 84.5%.

Apart from refining the models, we also analyze their *level-2* accuracy scores. It is intuitive to understand how *level-1* classification (wider, 3 nm range) may result in higher accuracy compared to *level-2* classification (narrower, 1 nm range). As depicted in Fig. 6(A), kNN again furnishes the best results of 73.8%, while RF and DT produce scores of 71.6% and 67.5%, respectively.

Since these results reflect the overall classification performance, averaged over all classes and subclasses, there is no information regarding the accuracy scores at specific classes or thicknesses. Hence, we next examine *level-1* accuracy and *level-2* accuracy by analyzing model performance at each class and thickness, respectively.

Figure 6(B) illustrates the accuracy of kNN at all classes, and as evident from it, we generally witness reduced accuracy scores at classes comprising higher thicknesses. While class 0 (1–3 nm) has the maximum accuracy of 98.6%, class 14 (43 and 44 nm) is associated with the minimum accuracy of 53.4%. There is a considerable reduction in accuracy of around 30%, which can mainly be attributed to the fact that class 14 encompasses only two subclasses (43–44 nm), unlike the other classes, which consist of three subclasses. Moreover, since the ALD process is terminated at 51.17 min, as soon as a film of 44 nm thickness is obtained, we have access to limited datapoints for this thickness. Next, we increase the granularity and examine kNN classification performance at each

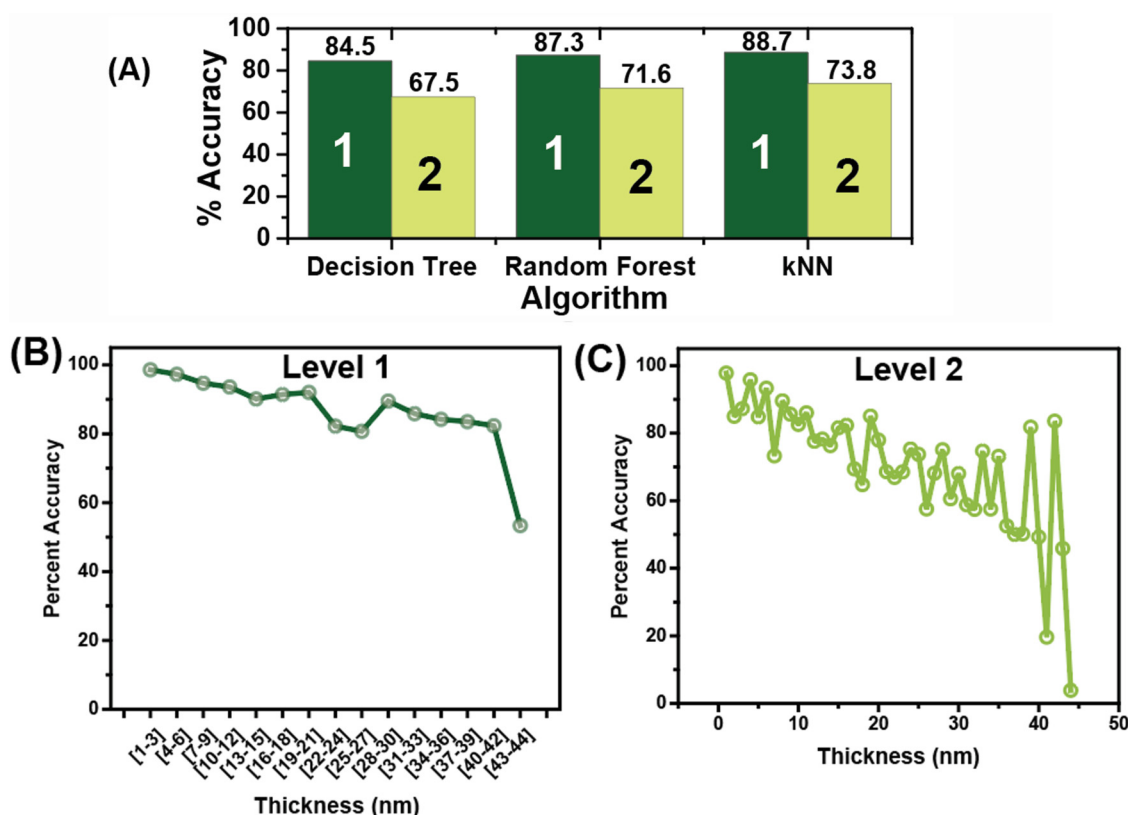


FIG. 6. (A) Accuracy results of the top three ML algorithms at classification levels-1 and -2 for collated ellipsometry data. Level-1 and level-2 accuracy depict the confidence of each algorithm to classify a particular *in situ* data point into one of the 15 main classes and 3 subclasses of the dataset, respectively. (B) Variation of level-1 accuracy with thickness for the kNN algorithm illustrates the accuracy of the kNN model at different bins of thickness values. (C) Variation of level-2 accuracy with thickness for the kNN model. A reduction in accuracy with an increase in thickness is observed for both levels-1 and -2.

thickness, as depicted in Fig. 6(C). The accuracy values of the model range between 97.8% for 1 nm and 3.9% for 44 nm. Although the general trend is negative, there are a few outliers (such as 39 and 42 nm) that produce high accuracy scores. The performance degradation at higher thicknesses may be attributed to the amplification of spectral features in Ψ and Δ [i.e., peaks and valleys in Figs. 3(A) and 3(B)], which hinder the learning stage of the models and degrade their classification performance.

C. Prediction accuracy is degraded upon increased randomized downsampling of SE data

The third case study describes the results of downsampling data, i.e., the removal of random Ψ and Δ values from the curated dataset. Here, we seek to understand how limited spectroscopic data might impact ML prediction accuracy. In the ultimate limit, where only one wavelength (as is the case for single wavelength ellipsometry) is present, might ML accurately predict film thickness? The results assume importance because limited data might indicate avenues for faster data acquisition, reduced hardware

complexity (i.e., costs), and, therefore, easier integration of SE for *in situ* monitoring of thin film deposition processes, in general.

The downsampling rate is varied between 0% and 90% and the performance of the best performing algorithm, the kNN algorithm, is evaluated on the downsampled dataset. Figure 7(A) depicts the variation of level-1 and level-2 accuracy at different downsampling rates. The level-1 accuracy is generally resilient to downsampling, varying between 88.7% at 0% and 85.7% at 85%, witnessing a maximum accuracy drop of 3%. The level-2 accuracy varies from 73.8% at 0% downsampling to 65.5% at 90% downsampling. This is an important result as it indicates that thickness prediction via ML can be effectively accurate to within ± 1.5 nm uncertainty (i.e., level-1) of a ZnO film ≤ 44 nm, even if 90% of the spectral data are eliminated.

On the other hand, Fig. 7(B) shows that downsampling induces a larger degradation in level-2 accuracy. This is equivalent to 8.4% drop in classification accuracy of the kNN model. However, the accuracy drops are relatively low at both levels until 50% downsampling. The performance degradation at level-1 and level-2 for 50% downsampling rate is 0.5% and 2%, respectively. These results are indicative of possible redundancy in the dataset with regard to Ψ and Δ samples.

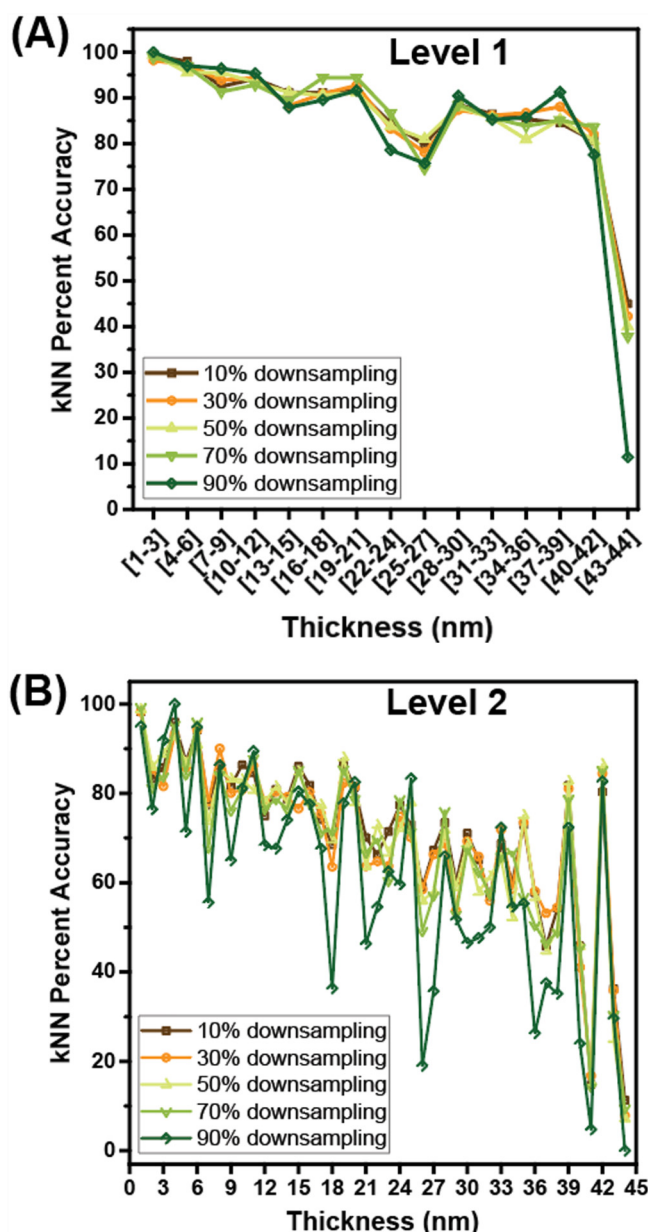


FIG. 7. (A) Variation of level-1 accuracy with thickness for select random downsampling rates illustrates the accuracy of the kNN model at different bins (3 nm each) of thickness values, which attenuates at higher thicknesses. (B) Variation of level-2 accuracy with thickness for the kNN algorithm at selected random downsampling rates which illustrates a negative trend, i.e., a reduction in accuracy with an increase in thickness.

We further study the variation of *level-1 accuracy* and *level-2 accuracy* with thickness at selected downsampling rates of 10%, 30%, 50%, 70%, and 90%. [Figure 8\(A\)](#) shows the *level-1 accuracy* at different classes (each bin with 3 nm thickness), and [Fig. 8\(B\)](#) shows

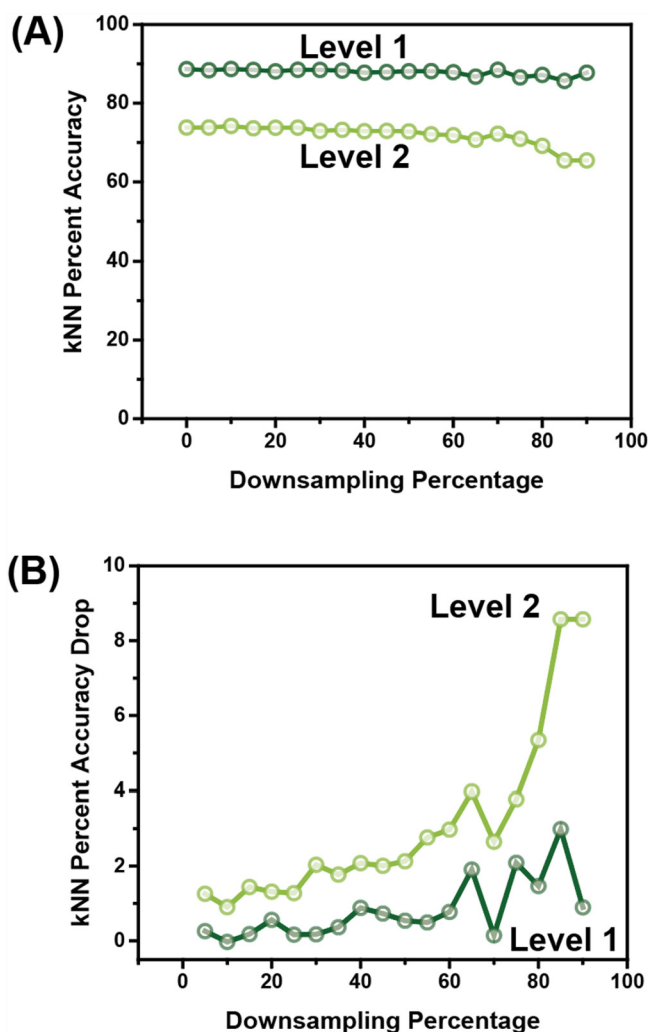


FIG. 8. (A) Accuracy results of the kNN algorithm at classification levels-1 and -2 for random downsampling of collated *in situ* ZnO ellipsometry data. Both level-1 and level-2 accuracies tend to degrade at higher downsampling rates. (B) Alternately, we show the variation of level-1 and level-2 accuracy drops of the kNN algorithm for random downsampling rates which illustrates degradation in accuracy for increased downsampling rate.

the *level-2 accuracy* at different thickness values. From these graphs, it is evident that the classification performance of kNN degrades at higher thicknesses, irrespective of the random downsampling rate. For example, with 90% downsampling of SE data, *level-1* prediction accuracy drops from 97% for a thickness in bin (4–6 nm) to 77.6% for a thickness in bin (40–42 nm). Thus, it is the actual physical thickness of the film predicted using ML that can make a difference to the predictive accuracy. This result bodes well for ML algorithms to predict the thickness of ultrathin films (≤ 10 nm) which is always a challenge for SE data analysis. Finally, it can be inferred that the dataset has redundancy that can be eliminated without significant loss in prediction accuracy. Hence, this motivates us to pursue

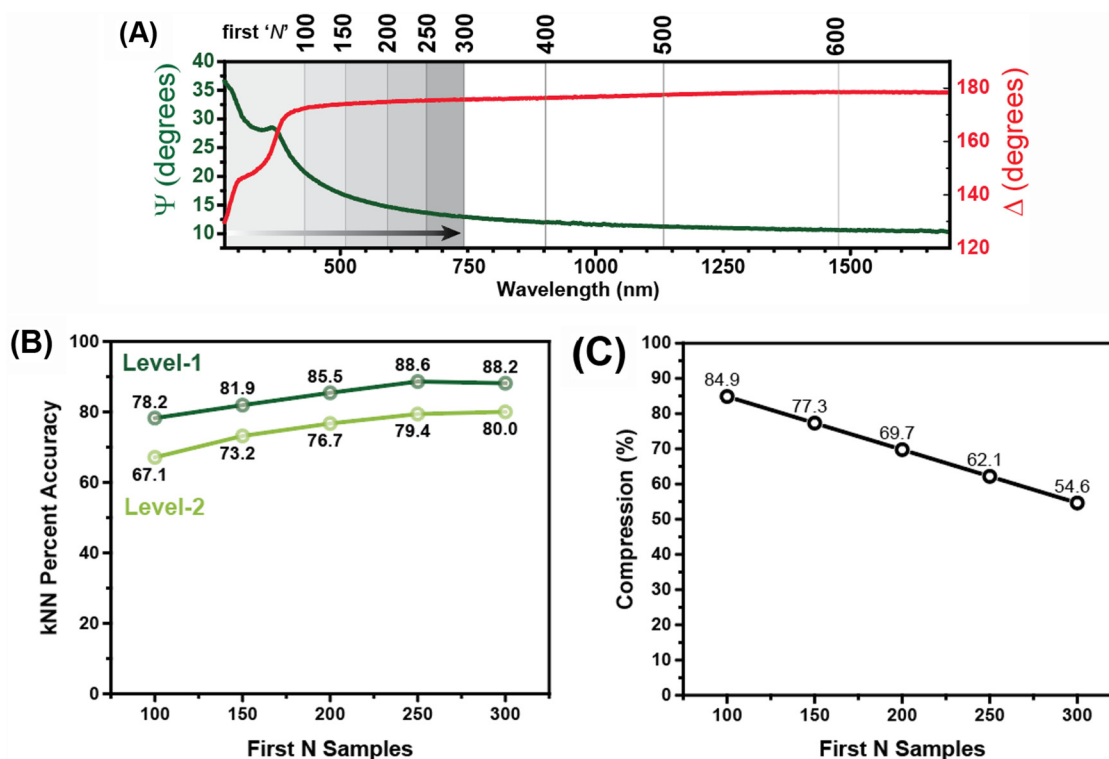


FIG. 9. (A) Ψ (left axis) and Δ (right axis) data showing the first N data points where $N = 100, 150, 200, 250$, and 300 . (B) Variation of kNN percent accuracy for level-1 and level-2 classifications as a function of selecting the first N wavelengths. (C) Variation of data compression with number of samples selected illustrating the increased fraction of data selected (i.e., lower compression) with increasing N .

selective downsampling and explore the connection, if any, between specific Ψ and Δ samples and classification accuracy.

D. Prediction accuracy is determined primarily by the first 300 data points on the SE spectra

The ultimate objective of an ML algorithm would be to determine thickness from minimal, perhaps few and discrete, wavelength based ellipsometry data. Therefore, the fourth case study involves selective (i.e., nonrandom) downsampling of Ψ and Δ values in the curated dataset. We select the first N wavelengths from each class and subclass, starting from 271 nm and up, while eliminating the rest. Hence, only the first N wavelengths from each thickness are selected, thereby sampling the remaining ones. This is shown in Fig. 9(A) which serves to remind the reader of the Ψ and Δ values of a single spectrum (i.e., related to a fixed thickness). Features such as peaks and valleys appear in the spectra for wavelengths ≤ 500 nm. It is important to note that the Ψ and Δ are experimentally measured raw data and no modeling effort has been expended to modify the data.

In this analysis, we consider N ranging from 100 to 300, in step sizes of 50. This corresponds to a spectral window from 271 to 427 nm for the first 100 and from 271 to 743 nm for first 300, in step sizes of 76 nm. From Fig. 9(A), an estimate can be made that

the first 300 data points still constitute a fraction (45.4%) of the total data collected.

This downsampled dataset is utilized to train and evaluate kNN, which, in the previous three case studies, has furnished the best results. Figure 9(B) shows the variation of level-1 and level-2 classification accuracy scores. While we obtain the lowest scores of 78.3% at level-1 and 67.1% at level-2 for the first 100 samples, improved classification performance is produced at higher values of N . The best level-1 accuracy and level-2 accuracy are 88.6% and 80%, which are furnished when the first 250 and 300 samples, respectively, constitute the dataset (opposed to 88.7% level-1 and 73.9% level-2 accuracies when no downsampling is performed, as shown in Fig. 6). This approach of selective downsampling provides results that exceed the baseline accuracy scores of the kNN model by 6.1% at level-2.

Another benefit of selective downsampling is illustrated in Fig. 9(C). Apart from bolstering the classification accuracy, selective downsampling furnishes data compression with regard to the overhead associated with training and testing datasets. Figure 9(C) portrays the variation of compression with increasing number of data samples. While we obtain savings of 84.9% for the first 100 samples, we obtain 62.1% and 54.6% compression for the first 250 and 300 samples, respectively. This coupled with performance enhancement of kNN is beneficial in our endeavor to optimize *in situ* SE from an efficiency, hardware, and cost point of view.

IV. CONCLUSIONS

In summary, we have analyzed *in situ* SE data obtained during the ALD of ZnO using an ML-based approach. The *in situ* SE data have the distinct advantage of providing a quasi-continuous variation of thickness within a single experiment, thus, providing a standalone dataset to train an ML algorithm. Polarization data consisting of Ψ and Δ versus wavelength (271–1688 nm) were recorded every 3 s interval for a total of 51.17 min. This resulted in 1113 pairs of Ψ and Δ curves. A total of 95% of the above data were prefit to a general oscillation optical model to extract thickness. A maximum thickness of 44 nm was recorded. These data were curated and used as the training set for the ML algorithm, where a *level-1* classification consisted of 15 bins of 3 nm each, and a *level-2* classification consisted of 44 bins of 1 nm each.

The following are the primary conclusions. First, five ML algorithms were evaluated, namely, kNN, random forest, decision tree, support vector machines, and logistic regression. Out of these, the kNN algorithm performed the best with an average prediction accuracy of 88.7% to within a ± 1.5 nm thickness interval, and an average prediction accuracy of 73.8% to within a ± 0.5 nm thickness interval. Second, the prediction accuracy degrades across classification bins as the thickness increases, i.e., thinner films have better prediction accuracies than thicker films. Third, removing data points via random downsampling of the Ψ and Δ curves by up to 90% can still result in high average prediction accuracies of 85.7%, thus, showing the redundancy of the SE data, *once* an ML algorithm is adequately trained. Finally, in contrast to random downsampling, featureless regions of the Ψ and Δ curves at longer wavelengths (or in a specific “spectral window”) can be eliminated while maintaining reasonably high prediction accuracies. The average prediction accuracy of the kNN algorithm remains robust at 88.2% when considering only the first 45.4% of the data (i.e., from 271 to 743 nm).

The above conclusions highlight the role an ML-based approach can play in bypassing modeling complexity, reducing computational overhead costs and, perhaps, easing hardware requirements for SE both as a standalone technique and for use as a powerful *in situ* technique for evaluating thin film deposition processes in semiconductor manufacturing. While this work lays the foundation of an ML-analysis of standard films, it is expected that in the future, ML-based algorithms will play an important role in bolstering the evaluation and analyses of SE data of compositionally complex films.

ACKNOWLEDGMENTS

The authors wish to thank the Semiconductor Research Corporation (SRC) for their support of this work under Award No. 3026.001. Helpful discussions with Greg Pribil and James Hilfiker from J.A. Woollam® are gratefully acknowledged. C.F. was supported by National Science Foundation (NSF) (Award No. 1908167).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

REFERENCES

- S. M. George, *Chem. Rev.* **110**, 111 (2010).
- T. Hatanpaa, M. Ritala, and M. Leskela, *Coord. Chem. Rev.* **257**, 3297 (2013).
- T. J. Knisley, L. C. Kalutarage, and C. H. Winter, *Coord. Chem. Rev.* **257**, 3222 (2013).
- H. Kim, H.-B.-R. Lee, and W. J. Maeng, *Thin Solid Films* **517**, 2563 (2009).
- C. Marichy, M. Bechelany, and N. Pinna, *Adv. Mater.* **24**, 1017 (2012).
- X. Meng, X.-Q. Yang, and X. Sun, *Adv. Mater.* **24**, 3589 (2012).
- E. B. Yousfi, J. Fouache, and D. Lincot, *Appl. Surf. Sci.* **153**, 223 (2000).
- J. W. Elam and S. M. George, *Chem. Mater.* **15**, 1020 (2003).
- A. Rahtu, T. Alaranta, and M. Ritala, *Langmuir* **17**, 6506 (2001).
- A. C. Dillon, A. W. Ott, J. D. Way, and S. M. George, *Surf. Sci.* **322**, 230 (1995).
- S. K. Park, R. Kanjolia, J. Anthis, R. Odedra, N. Boag, L. Wielunski, and Y. J. Chabal, *Chem. Mater.* **22**, 4867 (2010).
- K. Bernal-Ramos, M. J. Saly, R. K. Kanjolia, and Y. J. Chabal, *Chem. Mater.* **27**, 4943 (2015).
- K. B. Ramos, G. Clavel, C. Marichy, W. Cabrera, N. Pinna, and Y. J. Chabal, *Chem. Mater.* **25**, 1706 (2013).
- A. C. Kozen, A. J. Pearse, C. F. Lin, M. A. Schroeder, M. Noked, S. B. Lee, and G. W. Rubloff, *J. Phys. Chem. C* **118**, 27749 (2014).
- B. Brennan, X. Y. Qin, H. Dong, J. Kim, and R. M. Wallace, *Appl. Phys. Lett.* **101**, 211604 (2012).
- T. Kaufman-Osborn, E. A. Chagarov, S. W. Park, B. Sahu, S. Siddiqui, and A. C. Kummel, *Surf. Sci.* **630**, 273 (2014).
- T. Kaufman-Osborn, E. A. Chagarov, and A. C. Kummel, *J. Chem. Phys.* **140**, 204708 (2014).
- M. Schuisky, J. W. Elam, and S. M. George, *Appl. Phys. Lett.* **81**, 180 (2002).
- J.-S. Na, Q. Peng, G. Scarel, and G. N. Parsons, *Chem. Mater.* **21**, 5585 (2009).
- D. D. Fong, J. A. Eastman, S. K. Kim, T. T. Fister, M. J. Highland, P. M. Baldo, and P. H. Fuoss, *Appl. Phys. Lett.* **97**, 191904 (2010).
- L. Henn-Lecordier, W. Lei, M. Anderle, and G. W. Rubloff, *J. Vac. Sci. Technol. B* **25**, 130 (2007).
- J. M. Lowinsbury, J. A. Gladden, C. T. Campbell, I. S. Kim, and A. B. F. Martinson, *Chem. Mater.* **29**, 8566 (2017).
- E. Langeris, S. B. S. Heil, H. C. M. Knoop, W. Keuning, M. C. M. van de Sanden, and W. M. M. Kessels, *J. Phys. D: Appl. Phys.* **42**(7), 073001 (2009).
- E. Langeris, S. B. S. Heil, M. C. M. Van De Sanden, and W. M. M. Kessels, *J. Appl. Phys.* **100**, 023534 (2006).
- E. A. Irene, *In Situ Real-Time Characterization of Thin Films* (Wiley, New York, 2001).
- R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, *npj Comput. Mater.* **3**, 54 (2017).
- F. Lussier, V. Thibault, B. Charron, G. Q. Wallace, and J.-F. Masson, *Trends Anal. Chem.* **124**, 115796 (2020).
- A. A. Enders, N. M. North, C. M. Fensore, J. Velez-Alvarez, and H. C. Allen, *Anal. Chem.* **93**, 9711 (2021).
- W. B. Park, J. Chung, J. Jung, K. Sohn, S. P. Singh, M. Pyo, N. Shin, and K. S. Sohn, *IUCr* **4**, 486 (2017).
- J. Liu, D. Zhang, D. Yu, M. Ren, and J. Xu, *Light Sci. Appl.* **10**, 55 (2021).
- J. A. Woollam, *CompleteEASE Software Manual* (J. A. Woollam Co., Lincoln, USA, 2014).
- W. A. M. Harl and G. Tompkins, *Spectroscopic Ellipsometry and Reflectometry* (Wiley, New York, 1999).
- A. C. Kozen, A. J. Pearse, C.-F. Lin, M. Noked, and G. W. Rubloff, *Chem. Mater.* **27**, 5324 (2015).
- Z. Baji, Z. Labadi, Z. E. Horváth, G. Molnár, J. Volk, I. Bársony, and P. Barna, *Cryst. Growth Des.* **12**, 5615 (2012).
- I. H. Sarker, *SN Comput. Sci.* **2**, 160 (2021).
- E. Fix and J. L. Hodges, *Int. Stat. Rev.* **57**, 238 (1989).
- N. S. Altman, *Am. Stat.* **46**, 175 (1992).