1. Natalia PIEGAT, 2. Wojciech KIJASZEK, 3. Sergiusz PATELA, 4. Artur WYMYSŁOWSKI

ORCID: 1. 0009-0000-8832-7021 2. 0000-0002-2429-2701 3. 0000-0002-7007-0712 4. 0000-0003-4334-0972

OPEN ACCESS

DOI: 10.15199/48.2025.05.19

Influence of ellipsometric data type on SiO_2 thickness prediction accuracy for different machine learning algorithms

Wpływ typu danych elipsometrycznych na dokładność przewidywania grubości SiO_2 dla różnych algorytmów uczenia maszynowego

***Abstract**. The accurate measurements of dielectric thin film thickness is crucial for various applications in semiconductor manufacturing. This study investigates the influence of different types of ellipsometric data on the prediction accuracy of SiO_2 thickness using various machine learning algorithms. The performance of algorithms such as decision tree, k-nearest neighbours and random forest was compared when trained on different ellipsometric data types. The algorithms were trained on datasets for SiO_2 films with thickness ranging from 1 to 50 nanometers. The accuracy of the trained algorithms was verified on measurement results of prepared SiO_2/Si test structures. The analysis of the experiment results showed that the choice of ellipsometric data significantly impacts the prediction accuracy of the investigated algorithms.*

***Streszczenie.** Dokładne pomiary grubości cienkich warstw dielektrycznych mają kluczowe znaczenie dla różnych zastosowań w przemyśle półprzewodnikowym. W pracy zbadano wpływ różnych typów danych elipsometrycznych na dokładność predykcji grubości SiO_2 przy zastosowaniu różnych algorytmów uczenia maszynowego. Porównano wydajność algorytmów, takich jak drzewo decyzyjne, k-najbliższych sąsiadów i losowy las decyzyjny, po przeszkoleniu na różnych typach danych elipsometrycznych. Algorytmy zostały wytrenowane na zestawach danych dla warstw SiO_2 o grubości od 1 do 50 nanometrów. Dokładność wytrenowanych algorytmów została zweryfikowana na wynikach pomiarów przygotowanych struktur testowych SiO_2/Si. Analiza wyników eksperymentu wykazała, że wybór danych elipsometrycznych znacząco wpływa na dokładność przewidywania badanych algorytmów.*

Keywords: Spectroscopic ellipsometry, machine learning, decision tree, k-Nearest Neighbours, Random Forest
Słowa kluczowe: elipsometria spektroskopowa, uczenie maszynowe, drzewo decyzyjne, k-najbliższych sąsiadów, losowy las decyzyjny

Introduction

The accurate determination of thin film thickness is crucial in various scientific and industrial applications, including semiconductor manufacturing, optical coatings, and material science. Traditional methods for measuring thin film thickness, such as spectroscopic ellipsometry, provide high precision but often require complex data interpretation and modeling [1]. With the increasing popularity of machine learning (ML) techniques, there is a growing interest in application of advanced computational methods to enhance accuracy and simplify the prediction of thin film properties [2].

Spectroscopic Ellipsometry (SE) is a non-destructive optical technique applied for characterization of thin films. The SE system typically consists of a light source, a polarizer, a compensator, a sample stage, an analyser, and a detector (Fig. 1.). The light source emits a beam of polarized light, which is directed at the sample at a specific angle of incidence (AOI). The reflected light, which undergoes changes in its polarization state due to its interaction with the investigated material, is then analysed to extract information about the properties of the sample. The ellipsometric data is collected and analysed over a range of wavelengths. In the SE measurements, the change in polarization state is represented by the parameters Psi (Ψ) and Delta (Δ), which correspond to the amplitude ratio and phase difference between the p- and s-polarized light components, respectively (1).

$$\rho = \frac{R_p}{R_s} = \tan(\Psi)e^{i\Delta} \tag{1}$$

where: ρ – ellipsometric parameter, R_p,R_s – reflection coefficients for p- and s-polarized light.

The acquired data (Ψ and Δ) are fitted to a model that describes the sample's optical properties. This model typically includes parameters such as film thickness (d), refractive index (n) and extinction coefficient (k).

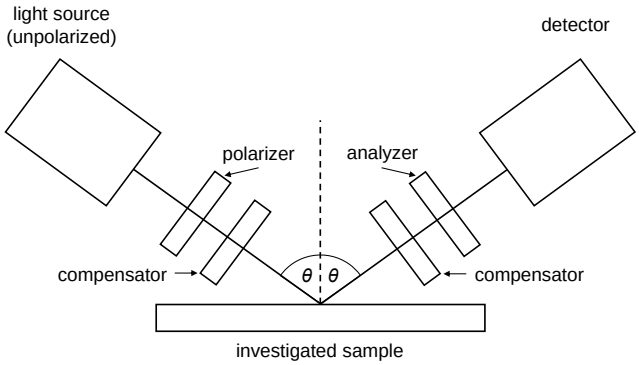

Fig. 1. Schematic diagram of ellipsometric measurement

Advanced models may also account for surface roughness, inhomogeneity and anisotropy [1]. The fitted model is validated by comparing the measured data with the model predictions. Fit quality parameters, such as the mean squared error (MSE), are used to assess the accuracy of the model.

Accurate measurements rely on the use of appropriate ellipsometric models, which describe how incident and reflected light interact with the material. These models are essential for interpreting the raw data and extracting meaningful parameters such as refractive index, extinction coefficient and film thickness. However, the process often requires the operator to make assumptions about initial parameters of the model based on their knowledge and experience. This subjective element introduces the potential for errors, as incorrect initial assumptions can lead to inaccurate results. Therefore, expertise in both the theoretical and practical aspects of ellipsometry is crucial for minimizing such errors and ensuring reliable measurements.

The application of machine learning in spectroscopic ellipsometry could enhance the accuracy and efficiency of model selection and result interpretation. Machine learning algorithms can be trained on extensive datasets of

ellipsometric measurements and corresponding model parameters, enabling them to identify patterns and correlations that might be overlooked by human operators. Therefore, machine learning can assist in selecting the most appropriate ellipsometric models, reducing the reliance on expertise of operators and minimizing the potential for human error. Additionally, machine learning can automate the analysis process, providing results and allowing operators to focus on more complex tasks. This integration of advanced computational techniques with traditional ellipsometry could lead to more reliable and reproducible measurements, ultimately advancing the field of thin film characterization.

In the literature on the application of machine learning in spectroscopic ellipsometry, researchers frequently employ and compare a variety of algorithms [2-4], including logistic regression, support vector machines, decision trees, random forests, k-nearest neighbours, but mostly artificial neural networks [5-12]. These comparisons are essential for identifying the most effective algorithm for specific tasks and datasets. Notably, many studies focus on datasets limited to a single type of ellipsometric data (i.e. Psi and Delta). Expanding the scope to include diverse forms of ellipsometric data could enhance the accuracy and versatility of machine learning, depending on the composition and structure of the analysed sample, leading to more comprehensive and reliable outcomes.

The ellipsometric data can be expressed not only as Ψ and Δ, but also as $\tan(\Psi)$ and $\cos(\Delta)$, real $\langle\varepsilon_1\rangle$ and imaginary $\langle\varepsilon_2\rangle$ part of the pseudodielectric function, real $\langle n\rangle$ and imaginary $\langle k\rangle$ part of the pseudo refractive index, real $\mathrm{Re}(\rho)$ and imaginary $\mathrm{Im}(\rho)$ part of ellipsometric parameter or Muller Matrix elements [1]. The influence of different ellipsometric data types on the accuracy of thin film thickness predictions using machine learning algorithms is explored in the paper. Various ellipsometric data types of thin silicon dioxide (SiO_2) films on silicon substrate are analysed to identify the optimal conditions under which the highest predictive performance of ML algorithms can be achieved.

The primary focus of the study was to investigate the impact of the type of ellipsometric data on the predictive accuracy of selected machine learning algorithms. Consequently, highly characterised test structures were selected for the purpose of verifying the operation of individual algorithms, namely thin SiO_2 films on silicon obtained through the dry thermal oxidation process. Given the markedly high degree of repeatability observed in the refractive index of SiO_2 films obtained through this technique, it was determined that the evaluation of the machine learning algorithms' performance and accuracy would be based on their capacity to predict SiO_2 film thickness.

The methodology section consists of the experimental section of the work and is accompanied by a detailed account of the data acquisition and preparation processes. The algorithms employed and the metrics utilized for assessing the precision of the results are outlined. The chapter dedicated to the presentation of results and discussion compares the accuracy of the obtained films thickness predictions and subsequently attempts to explain the observed relationships. The summary presents the most significant conclusions, accompanied with an overview of the plans for further research.

Methodology

In the work the V-VASE ellipsometer manufactured by J.A. Woollam Company with VB-400 control box, HS-190 monochromator and WVASE software were applied for the spectroscopic ellipsometry experiment. The numerical datasets of ellipsometric data: Ψ, Δ, $\tan(\Psi)$, $\cos(\Delta)$, $\langle\varepsilon_1\rangle$, $\langle\varepsilon_2\rangle$, $\langle n\rangle$, $\langle k\rangle$, $\mathrm{Re}(\rho)$ and $\mathrm{Im}(\rho)$ at different angles of incidence (65°, 70° and 75°) were generated for silicon dioxide thin film on a silicon substrate. The spectroscopic ellipsometric data was generated in a spectral range characteristic and often employed for SE thickness measurements, ranging from 300 to 1000 nanometres [1]. The dataset consisted of thermally grown 40 SiO_2 films with a different thickness ranging from 1 to 50 nanometres, representing the range of film thicknesses at which the interference colour of SiO_2 on silicon substrates and oscillations in the reflectance spectrum are not yet clearly visible, making it difficult to accurately determine the film thickness using methods other than SE. The 80% of the data is used for training and the remaining 20% is used for testing.

For validation of the trained ML algorithms a dataset of 13 measured silicon dioxide thin films grown on p-type silicon substrates of crystallographic orientation (100) in dry oxidation process (T = 1050°C). The reference samples were measured in spectral range from 300 to 1000 nm with resolution of 10 nm at incident angles of 65°, 70° and 75°. Considering the repeatability of the instrument, In order to determine the thickness of the SiO_2 films in the test structures as accurately as possible, the ellipsometric Ψ and Δ data was acquired 10 times for each wavelength and averaged during the measurement. The thickness of the SiO_2 films was estimated from an ellipsometric model in WVASE software. According to the standard deviation of the measurement parameters, the thickness was calculated with 0.1 nm uncertainty for 95% confidence limit.

In this study, three machine learning algorithms were examined for their suitability for the analysis of ellipsometric data. The accuracy of SiO_2 thickness predictions was determined for: decision tree (DT), k-nearest neighbours (kNN) and random forest (RF) algorithms. The algorithms were selected for analysing spectroscopic ellipsometry data due to their unique features.

Decision trees, widely used in the clinical literature, guide diagnostic and treatment decisions by recursively splitting datasets based on feature values, aiming to create homogeneous subsets. This process continues until further splitting no longer improves prediction accuracy. In decision tree terminology, each node is called a “branch,” each terminal node a “leaf,” and the starting node the “root” [13]. DTs are particularly useful for identifying important features and making decisions based on specific criteria, which can be beneficial for understanding complex relationships in spectroscopic data.

KNN is a simple, non-parametric algorithm that classifies data points based on the majority class among its k-nearest neighbours and is effective for pattern recognition [14]. The algorithm can be used for prediction of properties from spectroscopic ellipsometry data by comparing the data to known samples.

Though random forest, in case of our analysis seems to be the most promising, it was the most complex one due to its ensemble nature, relying on multiple decision trees, providing the most accurate prediction of nonlinear data [15]. In random forest regression, the final prediction is the average of all individual tree predictions. Interestingly, random forests are sometimes said to resemble deep network architectures, as they can distinguish between many regions of feature space [16]. The RF algorithm is designed to be less affected by noise and capable of handling large datasets, which makes it a suitable choice for analysing spectroscopic ellipsometry data with many variables.

To evaluate the performance of the algorithms, evaluation metrics such as coefficient of determination (R^2), mean squared error (MSE), mean absolute forecasting error (MAE) and symmetric mean absolute percentage forecasting error (SMAPE) were calculated and compared. The metrics are given in Table 1. and are often applied in evaluating the performance of forecasting models, each providing insights into the accuracy and reliability of predictions.

The R^2 statistic measures the proportion of variance in the dependent variable that can be predicted from the independent variable. It indicates how well the data points align with a given model. An R^2 value closer to 1 indicates a stronger model fit, whereas values closer to 0 suggest that the model does not align well with the data.

The MSE is a commonly used metrics that measures the average squared difference between the estimated values and the actual value. Two or more statistical models could be compared using their MSEs, as a potential measure of how well they explain a given set of observations. The main advantages of MSE is that it penalises larger errors more severely. A lower MSE indicates a more accurate fit of the model to the data.

The MAE is a measure of the average magnitude of the errors in a set of predictions, without consideration of their sign. This is particularly useful when the objective is to assess the performance of a given model. It is an easily interpretable and reliable method for assessing the error of predictions. A lower MAE value indicates a smaller error.

The SMAPE metric is useful for measuring accuracy in cases where the scale of the data varies. It's a percentage-based metric that helps to provide a standardised measure of error. The metric ranges from 0 to 2 (or 0-200%), with lower values indicating better model performance. It is particularly useful when comparing models across different datasets, as it is not dependent on scale.

Table 1. Performance metrics

Metrics	Definition	Equation
R^2	coefficient of determination	$1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2}$
MSE	mean squared error	$\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$
MAE	mean absolute forecasting error	$\frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i-y_i\rvert$
SMAPE	symmetric mean absolute percentage forecasting error	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert\hat{y}_i-y_i\rvert}{(\lvert y_i\rvert+\lvert\hat{y}_i\rvert)/2}$

Results and discussion

The case study compares the prediction accuracy of the applied ML algorithms as summarized in Table 2. In the case of SiO_2 thin films it follows that the type of ellipsometric data has observable influence on the prediction accuracy of the thickness of the investigated SiO_2 thin film depending on the ML algorithm.

Among the investigated ellipsometric data types, the best prediction accuracy for SiO_2 thin films has been achieved for the Ψ, $\tan(\Psi)$ and $\mathrm{Im}(\rho)$ data types, while the worst performance was observed for the $\langle\varepsilon_2\rangle$ based prediction. The silicon dioxide is a transparent, loss-less dielectric material with negligible surface roughness.

It could be hypothesised that the Ψ-based prediction yielded the most promising results, given that in the case of SiO_2, these parameters are the most susceptible to change and carry the most information with regard to variation in the thickness of the SiO_2 film.

The $\langle\varepsilon_2\rangle$ parameter is indicative of the loss of the entire structure. In relation to the SiO_2/Si structure, this parameter provides the least information about the SiO_2 film due to its loss-less nature in the investigated wavelength spectrum and is primarily indicative of the substrate parameters. This may be the cause of the lower level of accuracy in the prediction of film thickness.

It is possible that for other materials, i.e. light-absorbing, porous materials with rough surface, alternative ellipsometric parameters (such as Δ) could potentially offer a more accurate prediction of film thickness than Ψ.

From the regression graph of the predicted thickness vs measured thickness for Ψ-based prediction (Fig. 2a) it can be observed that the prediction is correct regardless of the thickness of the investigated film, while for the $\langle\varepsilon_2\rangle$-based prediction (Fig. 2b) the prediction of the film thickness deteriorated as the film thickness increased.

An examination of the probability density plot (Fig. 3) of the thickness prediction error (e_d) offers insights into the distribution and variability of the predictive models' performance. The distribution of the e_d for the Ψ-based prediction (Fig. 3a) is similar to a normal distribution for the kNN and RF algorithms. However, for the DT algorithm, the distribution is observed to broaden. The standard deviation of the error for all algorithms is less than 1 nm, and the average error value is close to 0. This demonstrates that the thickness of the investigated SiO_2 film was correctly predicted.

In the case of the $\langle\varepsilon_2\rangle$-based prediction (Fig. 3b) the e_d distribution for the DT algorithm most closely resembles a normal distribution among the investigated algorithms. For the kNN and RF algorithms, an additional error population can be observed, corresponding to an underestimation of the SiO_2 film thickness for thicker films.

Table 2. Comparison of models performance

algorithm	parameter	ellipsometric data type used for prediction									
		Ψ	Δ	tan(Ψ)	cos(Δ)	$\langle\varepsilon_1\rangle$	$\langle\varepsilon_2\rangle$	$\langle n\rangle$	$\langle k\rangle$	Re(ρ)	Im(ρ)
DT	R^2	0.9995	0.9995	0.9995	0.9995	0.9988	0.9995	0.9988	0.9995	0.9946	0.9995
	MSE	0.131	0.131	0.131	0.131	0.346	0.131	0.346	0.131	1.546	0.131
	MAE	0.31	0.31	0.31	0.31	0.49	0.31	0.49	0.31	1.03	0.31
	SMAPE	0.03	0.03	0.03	0.03	0.08	0.03	0.08	0.03	0.15	0.03
kNN	R^2	0.9997	0.9995	0.9997	0.9996	0.9995	0.9451	0.9991	0.9993	0.9935	0.9997
	MSE	0.079	0.139	0.079	0.117	0.147	15.8	0.265	0.193	1.866	0.079
	MAE	0.23	0.31	0.23	0.29	0.34	2.04	0.47	0.31	1.12	0.23
	SMAPE	0.02	0.03	0.02	0.03	0.04	0.06	0.07	0.03	0.17	0.02
RF	R^2	0.9999	0.9998	0.9999	0.9999	0.9995	0.9646	0.9989	0.9998	0.9969	0.9999
	MSE	0.028	0.045	0.031	0.043	0.139	10.2	0.309	0.070	0.904	0.029
	MAE	0.13	0.19	0.14	0.18	0.30	1.54	0.45	0.17	0.72	0.12
	SMAPE	0.01	0.02	0.01	0.01	0.05	0.05	0.08	0.01	0.12	0.02

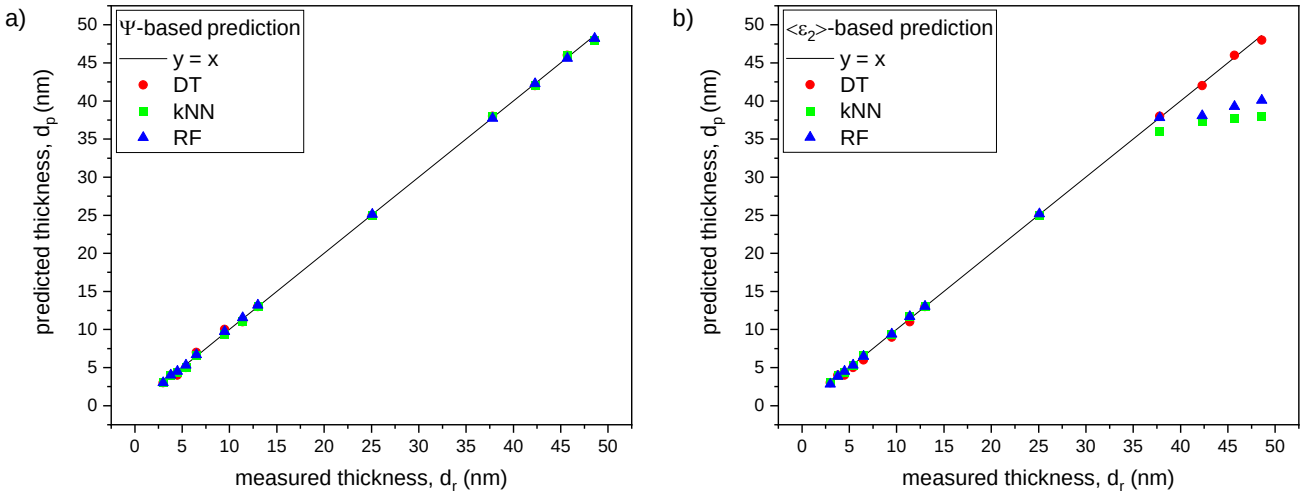

Fig. 2. Regression graphs of the predicted thickness vs measured thickness for: a) Ψ-based, b) <ε_2>-based prediction

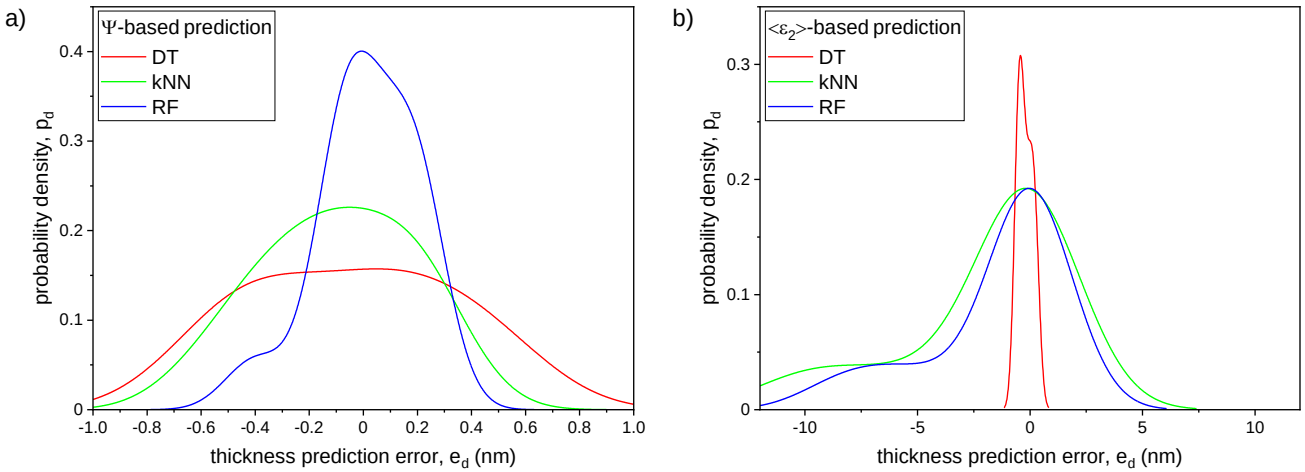

Fig. 3. Probability density plot of thickness prediction error for: a) Ψ-based, b) <ε_2>-based prediction

This observation can be explained by analysing the dispersion curves of individual ellipsometric data for films of varying thickness. The psi value, in the investigated thickness range, takes a certain value for a given wavelength only for a film of a given thickness (Fig. 4a). Moreover, if the changes in the shape of the dispersion curve are relatively minor, and the change in individual Ψ values with the thickness of the SiO_2 film is largely monotonic.

In the case of <ε_2>, the parameter assumes the same value repeatedly for a given wavelength when considering films of varying thickness (Fig. 4b). It is also important to note that the dispersion curve of the <ε_2> parameter undergoes a significant non-monotonic change in its course with the thickness of the SiO_2 film. Furthermore, depending on the considered spectral range, the parameter value undergoes a change in the opposite direction. The aforementioned factors may provide a supplementary explanation for the difficulties encountered in predicting the thickness of films from <ε_2> parameter.

It is noteworthy that the DT algorithm, despite exhibiting inferior performance compared to the RF and kNN algorithms, demonstrates satisfactory results for a broader range of ellipsometric data and across the full spectrum of film thicknesses. This can be explained by the fact that the DT algorithm approximates the analysed data to the model in its entirety, while the RF and kNN algorithms divide the data into smaller fragments. Consequently, the DT algorithm considers both the values and the shape of the dispersion curves of the recognised data, which makes it less prone to misidentifying film thickness.

A more precise conclusion can be reached by analysing the evaluation metrics calculated for each individual ML algorithm. The RF algorithm exhibited the highest accuracy, with an average layer thickness prediction accuracy of approximately 0.1 nanometres, representing an error of approximately 1%. The accuracy of the prediction of the thickness of SiO_2 layers grown by thermal dry oxidation of silicon substrates based on machine learning algorithms is slightly improved comparing to that presented in the literature on the application of similar algorithms for the prediction of the thickness of ZnO layers deposited on silicon substrates by the ALD technique [3]. The prediction accuracy obtained in this study is comparable to that achieved using neural networks [10], while more complex neural network yields superior results with SiO_2 thickness prediction error at the level of 0.1% [8].

Conslusions

In the paper, the influence of ellipsometric data type on SiO_2 thickness prediction accuracy was investigated using various machine learning algorithms. Three ML algorithms were evaluated, namely, decision tree, k-nearest neighbours and random forest with the following conclusions. Firstly, the best performance was achieved for the random forest algorithm with mean thickness prediction error of 0.1 nm. Secondly, the prediction accuracy depends on the thickness of the investigated film and the accuracy worsens for thicker films. Finally, the Ψ, tan(Ψ) and Im(ρ) data types yielded the most accurate results.

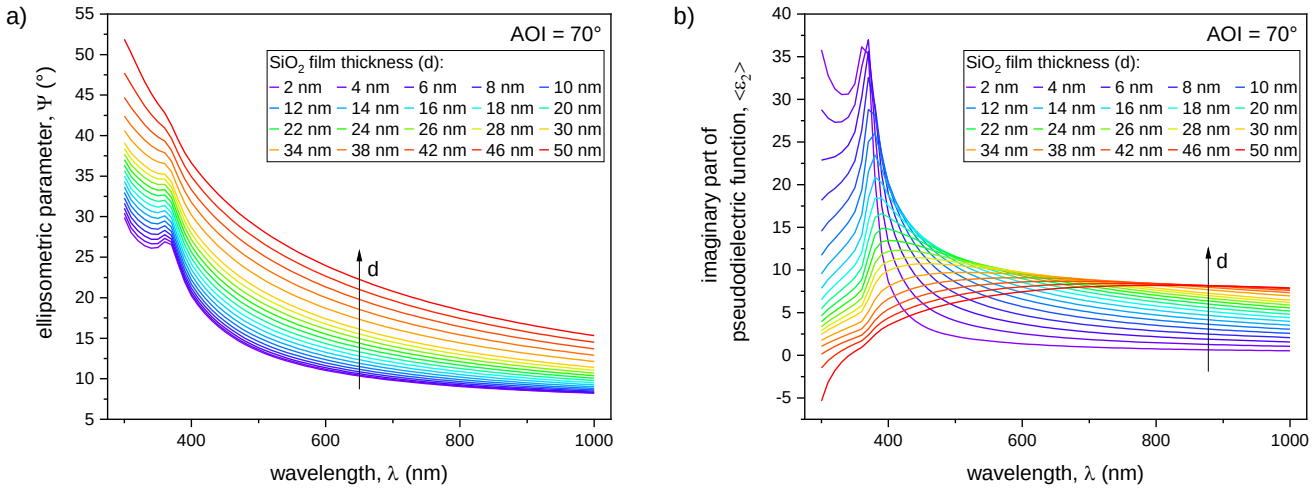

Fig. 4. Dispersive curves of ellipsometric parameters: a) Ψ b) $<\varepsilon_2>$ for different thickness of SiO_2 on Si substrate

The prediction errors for other data types were significantly higher, which can be contributed to a change in the shape of the dispersion curves of ellipsometric data, resulting from a change in the thickness of the investigated thin film. Moreover, for different types of materials than SiO_2, other ellipsometric parameters could yield better results (i.e. Δ for roughness prediction or $<\varepsilon_2>$ for absorbing materials. Consequently, the systematic analysis and curation of data sets tailored for training of the machine learning algorithm for specific materials could enhance the accuracy of predictive models.

Future research is planned to focus on exploring the applicability of these findings to other dielectric thin films (like Si_3N_4 or diamond-like carbon) and other materials, such as semiconductors and metals. Regression models may offer a deeper understanding of the relationship between the accuracy of thickness predictions and the methodologies employed in dataset curation. Additionally, this approach would enable the analysis of interactions among individual data, providing insights into the functioning and optimization of the model, as well as enhancing our understanding of the physical properties of the material. We aim to investigate these aspects in the subsequent phase of our research. However, appropriate transformation, standardisation, normalisation, balancing of the data set are required. The listed methods along with the optimisation of the hyperparameters (i.e. maximum depth, sample splitting, number of estimators, number of analysed features) of the algorithms for specific ellipsometric data types are required, which can lead to better accuracy of thickness and physical properties predictions of more complex structures.

The research was co-financed under the subsidy of the Wroclaw University of Science and Technology.

***Authors**: Natalia Piegat, Wrocław University of Science and Technology, Faculty of Electronics, Photonics and Microsystems, ul. Janiszewskiego 11/17, 50-372 Wrocław, E-mail: 268551@student.pwr.edu.pl; dr inż. Wojciech Kijaszek, Wrocław University of Science and Technology, Faculty of Electronics, Photonics and Microsystems, ul. Janiszewskiego 11/17, 50-372 Wrocław, E-mail: wojciech.kijaszek@pwr.edu.pl; dr hab. inż. Sergiusz Patela, Wrocław University of Science and Technology, Faculty of Electronics, Photonics and Microsystems, ul. Janiszewskiego 11/17, 50-372 Wrocław, E-mail: sergiusz.patela@pwr.edu.pl; prof. dr hab. inż. Artur Wymysłowski, Wrocław University of Science and Technology, Faculty of Electronics, Photonics and Microsystems, ul. Janiszewskiego 11/17, 50-372 Wrocław, E-mail: artur.wymyslowski@pwr.edu.pl*

REFERENCES

[1] Tompkins, H.G., Hilfiker, J.N., Spectroscopic ellipsometry: practical application to thin film characterization, *Momentum Press*, New York (2016)
[2] Mao Y., Wang L., Chen C., Yang Z., Wang J., Thickness Determination of Ultrathin 2D Materials Empowered by Machine Learning Algorithms, *Laser & Photonics Reviews*, 17 (2023), art. 2200357
[3] Arunachalam A., Berriel S.N., Feit C., Kumar U., Seal S., Basu K., Banerjee P., Machine learning approach to thickness prediction from in situ spectroscopic ellipsometry data for atomic layer deposition processes, *Journal of Vacuum Science & Technology A*, 40 (2022), issue 1, art. 012405
[4] Barkhordari A., Mashayekhi H.R., Amiri P., Özçelik S., Hanife F., Azizian-Kalandaragh Y., Performance of machine learning algorithms in spectroscopic ellipsometry data analysis of $ZnTiO_3$ nanocomposite, *Scientific Reports*, 14 (2024), art. 1617
[5] Urban F. K. , Barton D., Numerical ellipsometry: Artificial intelligence for rapid analysis of indium tin oxide films on silicon, *Journal of Vacuum Science & Technology A*, 42 (2024), issue 4, art. 043401
[6] Urban F.K., Barton D., Numerical ellipsometry: AI for real-time, in situ process control for absorbing films growing on unknown transparent substrates, *Journal of Vacuum Science & Technology A*, 42 (2024), issue 5, art. 053405
[7] Zhu P., Zhang D., Niu X., Liu J., Ren M., Xu J., A Lightweight Neural Network for Spectroscopic Ellipsometry Analysis, *Advanced Optical Materials*, 12 (2024), art. 2301381
[8] Li Y., Wu Y., Yu H., Takeuchi I., Jaramillo R., Deep Learning for Rapid Analysis of Spectroscopic Ellipsometry Data, *Advanced Photonics Research*, 2 (2024), issue 12, art. 2100147
[9] Kim I., Gwak S., Bae Y., Jo T., Optical spectrum augmentation for machine learning powered spectroscopic ellipsometry, *Optics Express*, 30 (2022), issue 10, pp. 16909-16920
[10] Jung J., Kim K., Choi J., Kim N., Chegal W., Cho Y., Kim Y., Geometric analysis algorithm based on a neural network with localized simulation data for nano-grating structure using Mueller matrix spectroscopic ellipsometry, *Optics Express*, 31 (2023), issue 26, pp. 44364-44374
[11] Liu J., Zhang D., Yu D., Ren M., Xu J., Machine learning powered ellipsometry, *Light: Science & Applications*, 10 (2021), art. 55
[12] Yang S., Chen X., Chen W., Hu J., Wang Y., Liu S., Liu S., Channeled spectroscopic ellipsometry enabled by physics-informed tandem untrained neural networks, *Measurement*, 235 (2024), art. 114940
[13] Rigatti S.J., Random Forest, *Journal of Insurance Medicine*, 47 (2017), issue 1, pp. 31-39
[14] Murty M.N., Avinash, M., Representation in Machine Learning, *Springer*, Singapore (2023)
[15] Breiman L., Random Forests, *Machine Learning*, 45 (2001), pp. 5-32
[16] Biau G., Scornet E., A random forest guided tour, *TEST*, 25 (2016), pp. 197-227