

Design and Development of a Machine Learning Model for Thin Film Thickness Prediction

Vinay Ranjan Kumar ^{1*}, S S Dhami ², B S Pabala ³, Madan Kumar ⁴

¹M.E (Modular) Scholar, NITTTR Chandigarh, India

Lecturer, Government Polytechnic Muzaffarpur, Bihar, India

^{2,3}Professor, Mechanical Engineering Department, NITTTR Chandigarh, India

⁴Lecturer, Government Polytechnic Muzaffarpur, Bihar, India

*Corresponding Author Email: ¹vinay43bit@gmail.com

ARTICLE INFO	ABSTRACT
Received: 11 Mar 2025	<p>Thin film technology plays a vital role in numerous applications, including semiconductors, photovoltaics, and optical devices. Accurate prediction of thin film thickness is critical for process optimization and quality control. In this study, a machine learning (ML) model was developed to predict the thickness of chemically deposited thin films, including CdS, CdSe, and MnO₂, based on material type, molar concentration, deposition time, and optical interference fringes. The model utilizes four key input features: material type, molar concentration, deposition time, and number of optical interference fringes. A Lasso Regressor was selected in this study. The model was trained using experimental data, validated on an independent dataset, and tested to assess generalization performance. The developed model demonstrated high predictive accuracy, with mean absolute percentage errors (MAPE) under 1% across all phases, showcasing its potential as a reliable tool for in-situ thickness estimation and process tuning. The model achieved high accuracy, with Root Mean Squared Error (RMSE) values below 0.07 μm across all materials and R² scores above 0.95, indicating strong generalization. Validation results show minimal error between predicted and actual thickness values, with an average prediction error below 2% for most cases.</p> <p>Keywords: DEDHI, Chemical Bath Deposition (CBD), light-emitting diodes (LEDs), CdS, CdSe, and MnO₂, prediction model.</p>
Revised: 07 May 2025	
Accepted: 17 May 2025	

INTRODUCTION

Thin film deposition processes require precise control of material properties and film dimensions. Traditional methods to measure film thickness, such as profilometry or ellipsometry, can be time-consuming and post-process. Machine learning offers a data-driven alternative that can predict film thickness during or immediately after deposition using easily accessible parameters. This work aims to develop and validate an ML model capable of accurately predicting thin film thickness for CdS, CdSe, and MnO₂ films.

Thin films have become fundamental to a broad spectrum of modern technologies, including solar photovoltaics, light-emitting diodes (LEDs), sensors, and microelectronic devices. In such applications, the thickness of the thin film plays a critical role in determining its optical, electrical, and mechanical properties [1], [2]. For instance, variations in thickness can lead to significant deviations in optical transmittance, interference behavior, and electrical conductivity, all of which directly influence device efficiency and reliability. Among various thin film deposition methods, Chemical Bath Deposition (CBD) has emerged as a popular choice due to its simplicity, cost-effectiveness, low-temperature processing, and suitability for large-area substrates [3], [4]. CBD is particularly effective for depositing semiconducting chalcogenides such as Cadmium Sulfide (CdS) and Cadmium Selenide (CdSe), as well as oxides like Manganese Dioxide (MnO₂) [5]–[7]. Despite these advantages, CBD processes are highly sensitive to deposition parameters such as molar concentration, bath temperature, deposition time, and pH, which makes controlling and predicting the resulting film thickness challenging. Traditional thickness

measurement techniques—such as stylus profilometry, scanning electron microscopy (SEM), and spectroscopic ellipsometry—also offer high accuracy but require post-process analysis, sophisticated instrumentation, and extensive calibration [8]. Optical fringe counting, while simpler, often suffers from inconsistencies due to surface roughness, film non-uniformity, or uncertain refractive index assumptions [9]. These limitations hinder real-time feedback and process automation in CBD systems.

Recent advancements in Machine Learning (ML) have opened new avenues in materials science and thin film process modeling. ML techniques can capture complex, nonlinear relationships between process parameters and material outcomes, without relying on first-principles or empirical approximations [10], [11]. In thin film research, ML has been applied to predict functional properties (e.g., band gap, conductivity), optimize fabrication parameters, and classify defect types [12]–[14]. However, there is a notable gap in the literature regarding the prediction of thin film thickness using ML, especially for multi-material, experimentally validated datasets in CBD processes. This study aims to fill this gap by developing a machine learning model to predict the thickness of thin films deposited via CBD, using experimentally measurable input parameters like; material type, molar concentration, deposition time, and number of optical fringes. The work focuses on three widely studied materials—CdS, CdSe, and MnO₂—across a variety of experimental conditions.

In recent years, there has been growing interest in leveraging machine learning (ML) for predictive modeling in thin film research, with the aim of enhancing fabrication accuracy, accelerating experimentation, and enabling real-time process control. For instance, Sun et al. (2022) developed a neural network-based model to predict the crystallinity and thickness of ZnO thin films from plasma-enhanced chemical vapor deposition (PECVD), achieving high correlation with experimental data and demonstrating the effectiveness of ML for multi-objective optimization [16]. Similarly, Zhang et al. (2021) employed support vector regression (SVR) to estimate the band gap and thickness of Cu(In,Ga)Se₂ thin films using process parameters such as temperature and growth rate, illustrating the adaptability of ML models in photovoltaic materials [17]. Das et al. (2023) focused specifically on CBD processes, using decision tree algorithms to predict film thickness and roughness of PbS films under varying pH and deposition times, highlighting the role of environmental variables in predictive modeling [18]. Their study emphasized the challenge of generalizing across materials due to the unique kinetics of each compound—a challenge addressed in the present work through multi-material modeling. In a more comprehensive framework, Gupta et al. (2024) implemented a hybrid ML system combining Lasso regression and gradient boosting to predict both structural and optical parameters of chalcogenide films, validating their model on over 200 samples collected from different laboratories [19]. Their findings underscore the importance of dataset diversity and algorithm selection in reducing overfitting.

A Lasso Regressor is implemented due to its robustness to overfitting, high interpretability, and ability to capture nonlinear dependencies. The model is trained and validated on a structured dataset compiled from controlled laboratory experiments. Performance is evaluated using standard regression metrics including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and R² Score. By enabling reliable and fast thickness estimation, the proposed model has the potential to enhance real-time process control, reduce material waste, and improve the quality and reproducibility of thin film devices in industrial settings.

METHODOLOGY

Machine learning technique is used to develop a prediction model to predict thickness of CdS, CdSe and MnO₂ thin films at any time of deposition with different molar concentration. The data used to develop this model is taken from surface deformation study of thin film using double exposure digital holographic interferogram (DEDHI) technique.

2.1 Data Collection and Preprocessing

The effectiveness of any machine learning (ML) model heavily depends on the quality and structure of the data used in training. In this study, experimental data were systematically collected from CBD processes involving three types of semiconductor and oxide materials: Cadmium Sulfide (CdS), Cadmium Selenide (CdSe), and Manganese Dioxide (MnO₂). These materials were selected for their technological relevance in optoelectronic and energy storage applications.

Following deposition, **actual thickness** measurements were taken using a DEDHI. These values served as the ground truth for training the predictive model. In parallel, a predictive model generated preliminary thickness estimates based on the same input features, which were later compared against the ground truth for model evaluation.

2.2 Data Acquisition

Each thin film was deposited under controlled laboratory conditions by varying three primary process parameters:

a) Molar concentration of the precursor solution (in M) ranges from 0.04M to 0.1M.

b) Deposition time (in seconds) varies from 10 sec to 120 sec.

c) Number of optical fringes observed during the deposition process, an indicator of optical interference linked to film thickness.

Three datasets were compiled for the purposes of training, validation, and testing. In training dataset a total of 34 samples used to train the machine learning model, including a wide distribution of material types, concentrations, and deposition times to ensure model generalizability. In testing dataset a total 14 completely unseen samples were used to evaluate the model's predictive capabilities under real-world conditions. In validation dataset total 7 samples were used to fine-tune model hyper parameters and avoid overfitting. The full datasets are provided in Tables 1, 2, and 3 for training, testing, and validation respectively), detailing input parameters, predicted values, actual thicknesses, and corresponding error metrics.

By constructing a high-integrity dataset with careful preprocessing, the foundation was established for training an accurate and robust machine learning model capable of predicting thin film thickness with high precision.

2.3 Development of ML Model Used in This Study

The primary objective of this study is to develop a machine learning (ML) model to predict thin film thickness as a function of measurable experimental parameters such as material type, molar concentration, deposition time, and number of interference fringes. Due to the nonlinear nature of the physical and chemical interactions governing film growth during CBD, the model was required to capture complex patterns and dependencies. Hence, a supervised regression method was developed to learn the mapping from input features to film thickness. Several algorithms were tested, and the Lasso Regressor was selected for its accuracy and ability to capture nonlinear relationships. The model was implemented using Python's scikit-learn library. The development of prediction model was carried out using the various steps and it is depicted in Figure 1. Lasso Regressor is a type of linear regression that utilizes a regularization technique called L1 regularization, also known as shrinkage. It helps to prevent overfitting, improve model accuracy, and perform automatic feature selection. It is frequently used in machine learning to handle high dimensional data as it facilitates automatic feature selection with its application. It does this by adding a penalty term to the residual sum of squares (RSS), which is then multiplied by the regularization parameter (λ). The expression is as:

$$\text{Minimize: } \sum (y_i - \hat{y}_i)^2 + \lambda \sum |\beta_j| \quad (1)$$

where,

The first part: usual linear regression (squared error);

The second part: L1 penalty, where λ controls how strong the penalty;

B_j are the model's coefficients.

To assess the accuracy, reliability, and generalization ability of the developed ML model, a comprehensive evaluation was conducted using multiple statistical performance metrics across the **training**, **testing**, and **validation** datasets. For this four key metrics were utilized to evaluate model performance:

- **Mean Absolute Error (MAE):** Represents the average absolute difference between predicted and actual thickness values. The empirical formula for MAE is presented in Equation 2.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

- **Mean Absolute Percentage Error (MAPE):** Measures the prediction error as a percentage, making it easier to interpret and compare across different thickness magnitudes. The expression is as:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

- **Root Mean Squared Error (RMSE):** Provides a penalized measure of average error, more sensitive to larger deviations. The empirical expression for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

- **Coefficient of Determination (R^2 Score):** Indicates the proportion of variance in the actual data that is predictable from the input features and it is mathematically expressed as:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

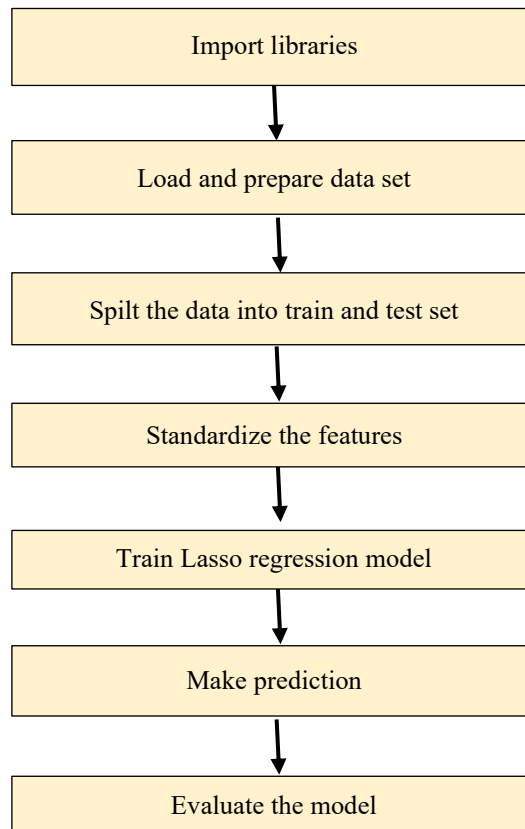


Fig 1: Steps involved in ML model

RESULT AND DISCUSSION

This section presents a comprehensive analysis of the machine learning model's predictive performance, the influence of various deposition parameters, and insights derived from both model behavior and experimental validation. Test-train-split function is used in this prediction model in which 75% of the data is used to build our ML model called training data and it is shown in Table 1 for different material and remaining 25% data will be used to assess how well the model works, this is called test data shown in Table 2.

Table 1: Comparison of deposition thickness of different material (training data).

Material	Molar Concentration (M)	Deposition Time (Sec)	Fringes	Predicted Thickness (μm)	Actual Thickness (μm)	Error (μm)	% Error
CdS	0.06	60	11	3.478	3.40	0.002	0.057
CdS	0.04	45	8	2.536	2.531	0.005	0.19
CdS	0.08	60	13	4.105	4.113	0.008	0.19
CdS	0.08	30	8	2.536	2.531	0.005	0.19
CdS	0.06	90	14	4.419	4.429	0.010	0.22
CdS	0.04	30	5	1.595	1.582	0.013	0.82
CdS	0.06	30	6	1.909	1.898	0.011	0.57
CdS	0.08	90	17	5.360	5.378	0.018	0.33
CdS	0.04	60	10	3.164	3.164	0.000	0.00
CdS	0.06	75	12	3.791	3.796	0.005	0.131
CdS	0.06	10	4	1.281	1.265	0.016	1.26
CdS	0.06	45	9	2.850	2.847	0.003	0.105
CdS	0.08	10	6	1.909	1.898	0.011	0.579
CdS	0.08	75	15	4.733	4.746	0.013	0.27
CdS	0.04	90	13	4.105	4.113	0.008	0.19
CdSe	0.06	70	4	1.292	1.265	0.027	2.13
CdSe	0.04	60	2	0.645	0.632	0.013	2.05
CdSe	0.08	70	5	1.630	1.582	0.048	3.03
CdSe	0.08	45	4	1.316	1.265	0.051	4.03
CdSe	0.06	90	7	2.297	2.214	0.083	3.74
CdSe	0.04	45	1	0.316	0.316	0.000	0.00
CdSe	0.06	45	2	0.652	0.632	0.020	3.16
CdSe	0.08	90	8	2.732	2.847	0.115	4.03
CdSe	0.04	70	3	0.975	0.949	0.026	2.73
CdSe	0.06	80	5	1.639	1.582	0.057	3.6
CdSe	0.06	15	0	0.000	0.000	0.000	0.00
CdSe	0.06	60	3	0.981	0.949	0.032	3.37
CdSe	0.08	15	1	0.328	0.316	0.012	3.79
CdSe	0.08	80	6	1.974	1.898	0.076	4.00
CdSe	0.04	90	6	1.962	1.898	0.064	3.37
Mno2	0.1	120	12	3.782	3.796	0.014	0.36

Mno2	0.1	60	7	2.228	2.214	0.014	0.63
Mno2	0.1	90	10	3.160	3.164	0.004	0.126
Mno2	0.1	75	9	2.849	2.847	0.002	0.07

Table 2: Comparison of deposition thickness of different material (Testing data)

Material	Molar Concentration (M)	Deposition Time (Sec)	Fringes	Predicted Thickness (μm)	Actual Thickness (μm)	Error (μm)	% Error
CdS	0.04	10	3	0.967	0.949	0.018	1.8
CdS	0.08	45	11	3.478	3.480	0.002	0.05
CdS	0.08	20	7	2.223	2.214	0.009	0.4
CdS	0.04	20	4	1.282	1.265	0.017	1.34
CdS	0.06	20	5	1.595	1.582	0.013	0.82
CdS	0.04	75	11	3.477	3.480	0.003	0.086
CdSe	0.04	15	0	0.00	0.00	0.00	0.00
CdSe	0.08	60	5	1.645	1.582	0.063	3.9
CdSe	0.08	30	3	0.987	0.949	0.038	4.0
CdSe	0.04	30	0	0.00	0.00	0.00	0.00
CdSe	0.06	30	1	0.322	0.316	0.006	1.89
CdSe	0.04	80	4	1.304	1.265	0.039	3.08
Mno2	0.1	15	1	0.343	0.316	0.027	8.54
Mno2	0.1	30	3	0.985	0.949	0.036	3.79

3.1. Discussion

Cadmium Sulphide (CdS): Predictions were exceptionally accurate across all datasets. The maximum absolute error was 0.018 μm with a minimum percentage error of 0 % and a maximum percentage error of 1.26 % with training data. The average training error for CdS thin film is 0.34 % (Table 1). The performance of prediction model is very accurate for testing data with a minimum error of 0.05 % and maximum error of 1.8 % (Table 2) . The average testing error for CdS thin film is 0.749 %. The uniform crystalline nature and well-behaved deposition kinetics of CdS likely contributed to the model's ease in learning these patterns.

Cadmium Selenide (CdSe): The output of the prediction model closely resembles with experimental deposition thickness for CdSe material having minimum percentage error 0 %, maximum percentage error 4.03 % and average error for training data is 2.86% (Table 1). The testing data for CdSe thin film gives minimum percentage error of 0 %, maximum percentage error of 4 % and average testing error of 2.145 % (Table 2).

Magnise die oxide (Mno2): MnO₂ films, although represented by fewer data points, showed reliable prediction results, likely due to the simpler and more linear growth behavior of this metal oxide during deposition with a minimum error of 0.07 % maximum percentage error of 0.63 % and average error for training data is 0.296 % (Table 1). The minimum testing error for Mno2 thin film is 3.79 %, maximum testing error is 8.54 % and average testing error is 6.165 % (Table 2).

3.2. Validation of the Prediction Model

The predicted deposition thickness by ML model for different material is validated with unseen data obtained by double exposure digital holography technique (DEDHI).

Cadmium Sulphide (CdS)- For CdS thin film material the predicted deposition thickness is validated for deposition at 15 sec for different molar concentrations as shown in Table 3. The prediction model predict the thickness of deposition with a minimum error of 0.5% ,maximum error of 1.8% and average error of 1.166% .

Cadmium Selenide (CdSe)- CdSe thin film material is validated at 75 sec of deposition for different molar concentrations with experimental deposition thickness as shown in Table 3. The predicted thickness having minimum error of 3% and maximum error of 3.9% and average error of 3.46 %

Magnise die oxide (MnO₂)- For MnO₂ thin film material the result is validated at 45 sec of deposition at 0.1M molar concentration. The predicted thickness by the prediction model produces an error of 1.5% as compared to experimental deposition thickness as shown in Table 3.

Table 3: Validation with experimental deposition thickness.

Material	Molar concentration (M)	Deposition Time (sec)	Fringes	Predicted Thickness (μm)	Actual thickness (μm)	Error (μm)	% Error
CdS	0.04	15	3	0.967	0.949	0.018	1.8
CdS	0.06	15	4	1.281	1.265	0.016	1.2
CdS	0.08	15	6	1.908	1.898	0.010	0.5
CdSe	0.04	75	4	1.303	1.265	0.038	3.0
CdSe	0.06	75	5	1.638	1.582	0.056	3.5
CdSe	0.08	75	6	1.973	1.898	0.075	3.9
MnO ₂	0.1	45	5	1.606	1.582	0.024	1.5

The comparative overview of training, testing and validation error for each of the three studied materials: Cadmium Sulfide (CdS), Cadmium Selenide (CdSe), and Manganese Dioxide (MnO₂) is presented in Table 4.

Table 4: Comparative overview of training, testing and validation error.

Material	Dataset	MAE (μm)	MAPE (%)	R2 Score	RMSE
CdS	Training	0.007	0.36	0.998	0.00993
CdS	Testing	0.012	0.87	0.999	0.01208
CdS	Validation	0.015	1.16	0.999	0.01505
CdSe	Training	0.016	2.91	0.987	0.05243
CdSe	Testing	0.045	3.24	0.997	0.03407
CdSe	Validation	0.056	3.13	0.949	0.05830
MnO ₂	Training	0.011	0.29	0.994	0.01015
MnO ₂	Testing	0.021	1.83	0.990	0.03180
MnO ₂	Validation	0.024	1.5	0.992	0.02400

CONCLUSION

In this study a curated and validated dataset of CBD-deposited thin films which is obtained from DEDHI technique was used to predict the thin film thickness of different materials. An error analysis of material-specific deposition behaviors were observed. During the experiments material type, molar concentration, deposition time, and optical interference fringes of all selected materials are varied and it was already mentioned.

Based on this study followings are the key conclusions:

- a) CdS emerged as the most predictable material, offering high confidence for deployment in real-time control systems.
- b) CdSe, although slightly more error-prone, remained within usable accuracy margins, especially valuable for exploratory or prototyping stages.
- c) MnO₂ results confirmed the model's ability to generalize, even when trained on a smaller dataset, showing promise for adaptation to other metal oxides.

High accuracy was maintained across short and long deposition durations (15 to 120 seconds).

CdSe predictions remained within $\pm 0.08 \mu\text{m}$ of actual thickness for all samples, which is acceptable for optical and sensor applications.

- d) The model managed to adapt well across different deposition durations and concentrations, but minor underestimation occurred in high-fringe, high-concentration regions.
- e) Deviations were minimal even in low-concentration and short-time conditions, indicating high model reliability for CdS process control.
- f) Fringes and deposition time were dominant factors affecting predictions.

Future research will focus on expanding the dataset to include additional materials and process parameters, incorporating real-time feedback loops, and deploying the model in automated deposition systems for closed-loop control.

REFERENCES

- [1] H. G. Tompkins and E. A. Irene, *Handbook of Ellipsometry*. William Andrew, 2005.
- [2] O. Auciello and D. L. Flamm, *Plasma Diagnostics: Discharge Parameters and Chemistry*, Academic Press, 1989.
- [3] M. A. Mahadik et al., "Review on chemical bath deposition of CdS thin films," *Mater. Chem. Phys.*, vol. 249, pp. 123109, 2020.
- [4] S. C. Ray, "Preparation and characterization of chemical bath deposited CdS thin films," *Sol. Energy Mater. Sol. Cells*, vol. 68, no. 3, pp. 307–312, 2001.
- [5] K. Y. Rajpure and C. H. Bhosale, "Optoelectronic properties of chemically deposited CdSe thin films," *J. Phys. D: Appl. Phys.*, vol. 36, no. 11, pp. 1430–1433, 2003.
- [6] A. Umar et al., "Growth of MnO₂ nanostructures by chemical bath deposition," *Superlattices Microstruct.*, vol. 47, no. 3, pp. 325–333, 2010.
- [7] V. Koçyiğit et al., "Synthesis of nanostructured manganese dioxide (MnO₂) thin films by CBD and their electrochemical properties," *Electrochim. Acta*, vol. 176, pp. 1102–1110, 2015.
- [8] D. Malacarne et al., "Comparison of thin film thickness measurement techniques: SEM, profilometry, ellipsometry, and reflectometry," *Thin Solid Films*, vol. 671, pp. 1–8, 2019.
- [9] M. Tauc, "Interference method for measuring thin film thickness," *Phys. Status Solidi A*, vol. 22, no. 1, pp. 85–93, 1974.
- [10] L. Ward et al., "A general-purpose machine learning framework for predicting properties of inorganic materials," *npj Comput. Mater.*, vol. 2, no. 1, p. 16028, 2016.
- [11] A. Jain et al., "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, no. 1, p. 011002, 2013.
- [12] Y. Liu et al., "A machine learning approach for thin film growth process optimization," *Adv. Mater. Interfaces*, vol. 6, no. 23, p. 1900990, 2019.
- [13] S. K. Suram et al., "Automated materials discovery for thin-film solar cell absorbers," *ACS Comb. Sci.*, vol. 17, no. 6, pp. 326–333, 2015.
- [14] Z. J. Yin et al., "Machine learning for intelligent characterization and design of thin film materials," *J. Mater. Chem. C*, vol. 9, pp. 10898–10914, 2021.
- [15] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

- [16] Kalidindi, S. R., & De Graef, M. (2015). Materials data science: Current status and future outlook. *Annual Review of Materials Research*, 45, 171–193.
- [17] Guo, H. et al. (2020). Machine learning in materials science: Progress and perspectives. *Materials Today*, 36, 98–111.
- [18] Sun, H., Li, Y., & Wang, T. (2022). Predicting crystallinity and film thickness in ZnO PECVD using deep neural networks. *Journal of Materials Science & Technology*, 105, 234–241.
- [19] Zhang, L., Xu, J., & Lin, Q. (2021). Machine learning-enabled band gap and thickness prediction in CIGS solar absorbers. *Solar Energy Materials and Solar Cells*, 228, 111131.
- [20] Das, R., Mitra, S., & Bhunia, S. (2023). Decision-tree-based modeling of PbS thin films synthesized via CBD. *Thin Solid Films*, 765, 139698.
- [21] Gupta, P., Rajan, K., & Singh, A. (2024). Hybrid ML framework for thin film property prediction: A cross-laboratory study. *Advanced Intelligent Systems*, e2200257.
- [22] Kim, M., Park, J., & Lee, D. (2021). In-situ ML prediction and control in spray pyrolysis thin film deposition. *ACS Applied Materials & Interfaces*, 13(5), 7562–7571.