Supplementary Information

Controlled growth of high-quality SnSe nanoplates assisted by machine learning

Huijia Luo[a], Wenwu Pan[a], Junliang Liu[a,b], Han Wang[a], Songqing Zhang[a], Yongling Ren[a], Cailei Yuan[c], and Wen Lei[a,*]

a. Department of Electrical, Electronic and Computer Engineering, The University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia.

b. Department of Electronic Engineering, School of IOT engineering, Jiangnan University, Wuxi 214122, China.

c. Jiangxi Key Laboratory of Nanomaterials and Sensors, School of Physics, Communication and Electronics, Jiangxi Normal University, 99 Ziyang Avenue, Nanchang 330022, China.

* Corresponding author. *E-mail address:* wen.lei@uwa.edu.au (W. Lei).

Growth process of SnSe NPs

The SnSe NPs were grown *via* a catalyst-free chemical vapour deposition (CVD) process in a single-zone horizontal tube furnace (Linderberg/Blue M). 15 mg of commercial SnSe powder (purity 99.99%) was placed in a quartz boat at the centre of the heating zone as the precursor material for evaporation, while seven pieces of 300 nm SiO_2/Si substrates (~ 0.5 cm × 0.5 cm in size) were placed in the downstream area (~ 11 cm - 15 cm from the precursor material). Before heating, the tube was vacuumed by the pump and purged with ultra-high-purity Ar carrier gas flow (100 sccm) for 10 mins to remove any air residue in the tube. Next, the CVD system works according to the designed growth conditions. For example, to obtain the largest SnSe NPs, the furnace was heated to 612 °C at a ramp speed of 25 °C/min without Ar carrier gas flow, after which the furnace was maintained at 612 °C for 48 mins with 88 sccm Ar gas flow under 4.9 Torr pressure to grow SnSe NPs. Once the reaction finished, the furnace was naturally cooled down to room temperature without Ar gas flow.

Machine learning algorithm

1. Gaussian Process Regression (GPR)

GPR is a non-parametric, probabilistic regression technique that leverages Gaussian processes to model the relationships between input variables and a continuous output variable. In GPR, a Gaussian process is used to define a distribution over functions. This allows the model to represent a range of possible functions that could describe the data. Suppose x is the input vector, $f(x)$ is the regression function, the mean and covariance of $f(x)$ is described as Equation (1) and (2), respectively:

$$E(f(x)) = m(x) \#(1)$$

$$Cov\left[f(x), f\left(x^{'}\right)\right] = E\left\{\left[f(x) - m(x)\right]\left[f\left(x^{'}\right) - m\left(x^{'}\right)\right]\right\} = k\left(x, x^{'}\right) \#(2)$$

then the Gaussian Process is written as:

$$f(x) \sim GP\left(m(x), k\left(x, x^{'}\right)\right) \#(3)$$

The GPR model with Gaussian noise can be formulated as follows:

$$y = f(x) + \varepsilon \#(4)$$

where y is the prediction, ε is the noisy or mismatch values that is assumed to follow the Gaussian distribution $\left(\varepsilon \sim N\left(0, \sigma_n^2\right)\right)$. Therefore, a GPR model is a probabilistic model in which a latent variable is

introduced for each observation, contributing to its nonparametric nature. The illustration GPR is given in **Fig. S1**.

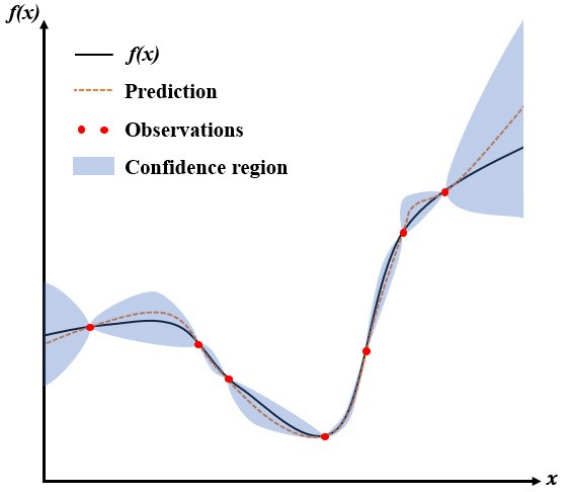

Fig. S1 Illustration of a GPR.

2. Support vector regression (SVR)

SVR is a classic implementation of support vector machines in regression problems. It operates on the principle of finding a hyperplane in a high-dimensional space that best fits the training data within a predefined margin of tolerance. Suppose a dataset of n points is given as:

$$\{(x_1,y_1),\ldots,(x_i,y_i)\}\#(5)$$

where x_i is a real input feature vector and y_i is the regression value. The regression function can be described as：

$$f(x) = w \cdot \varphi(x) + b\#(6)$$

where w is the weight factor, $\varphi(x)$ is the mapping function of inputs in high dimensional space and b is the bias. If the deviation between y_i and $f(x_i)$ is less than ε (the largest tolerance error), then the $f(x_i)$ is said to be found. Thus, a function $f(x)$ is desired that has most ε deviation from y_i, while being as flat as possible. An ε-insensitive loss function is introduced to describe the degree of deviation:

$$L(x,y,f) = \begin{cases} |y_i - f(x_i)| - \varepsilon & if\ |y_i - f(x_i)| \geq \varepsilon \\ 0 & otherwise \end{cases} \#(7)$$

This function indicates that training points will not be penalized if they are within the ε-tube range. In order to achieve the flatness of the desired regressor, the square of the norm of w should be minimized. The problem is described as follows:

$$minR(w) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} L(x,y,f) \#(8)$$

Slack variables δ_i and δ_i^* are introduced to cope with infeasible constraints. The above problem can be formulated into the following optimization function:

$$\min R(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)$$

$$s.t.\begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i \\ f(x_i) - y_i \le \varepsilon + \xi_i^* \#(9) \\ \xi_i, \xi_i \ge 0 \end{cases}$$

where C is the penalty parameter to determine the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger than ε are tolerated. **Fig. S2** depicts the basic principle of non-linear SVR to address problems with constraints, Lagrange multipliers can be used as follows:

$$L := \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{n}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right)$$
$$-\sum_{i=1}^{n}\alpha_i\left(\varepsilon + \xi_i - y_i + w\cdot\varphi(x_i) + b\right) - \sum_{i=1}^{n}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - w\cdot\varphi(x_i) - b\right) \#(10)$$

where L is the Lagrangian, $\alpha_i \ge 0$, $\alpha_i^* \ge 0$, $\eta_i \ge 0$, and $\eta_i^* \ge 0$ are Lagrange multipliers. When the constraint functions have strong duality, and the objective function is differentiable, Karush-Kuhn-Tucker conditions must be satisfied for each pair of the primal and dual optimal points as follows:

$$\begin{cases} \alpha_i\left(\varepsilon + \xi_i - y_i + w\cdot\varphi(x_i) + b\right) = 0 \\ \alpha_i^*\left(\varepsilon + \xi_i^* + y_i - w\cdot\varphi(x_i) - b\right) = 0 \\ (C - \alpha_i)\xi_i = 0 \\ \left(C - \alpha_i^*\right)\xi_i^* = 0 \end{cases}$$

By solving the above equations, the Lagrange dual problem can be derived as follows:

$$max\left(-\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)x_i^T x_j - \varepsilon\sum_{i=1}^{n}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{n} y_i\left(\alpha_i - \alpha_i^*\right)\right)$$

$$s.t.\sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right) = 0, and\ \alpha_i, \alpha_i^* \in [0, C] \#(11)$$

$$w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\varphi(x_i)$$

The weight vector can be obtained as , and therefore the regression function is:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\varphi(x_i)x + b \#(12)$$

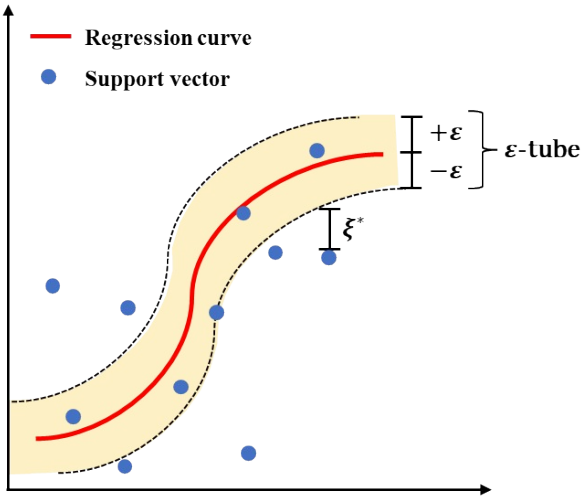

Fig. S2 Example of an SVR with an e-tube.

3. Regression tree (RT)

The construction of a RT involves recursively splitting the data based on the input features to minimize the variance of output variables within each split. The process starts at the root of the tree and divides the data into two subsets using the feature and threshold that result in the largest variance reduction. This division process repeats until a predefined stopping criterion is met. **Fig. S3** depicts an example of RT method by giving two input variables, X and Y, for a regression task. This splitting continues in a recursive manner on each subset until a stopping criterion is met, which could be a minimum number of observations in a node, a maximum tree depth, or a minimum improvement in variance reduction.

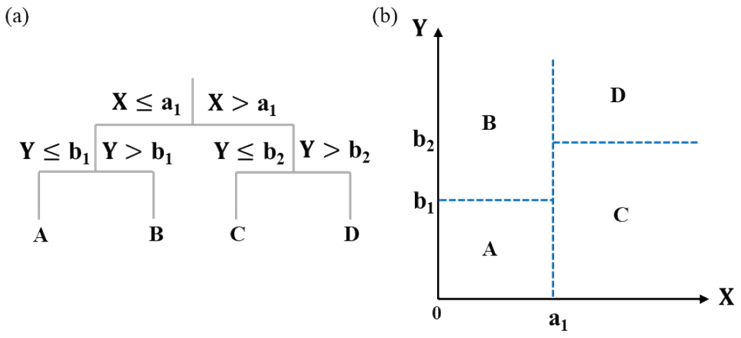

Fig. S3 Example of an RT.

4. Random forest (RF)

RF regression is an ensemble learning method that operates by constructing multiple RTs in the training process and the final output is the mean prediction of the individual trees. It extends the concept of a single RT to a forest of trees, aiming to reduce the variance and improve the accuracy of predictions for continuous target variables by referring to a bagging method, thereby overcoming some of the limitations of individual RT, such as overfitting. Assume that n samples are randomly collected from S_n with a selective probability of $1/n$ for each sample. These randomly selected n samples are called a bootstrap sample θ, where θ is an independently distributed vector. Assume that q bootstrap samples ($S_n^{\theta_1}$,…, $S_n^{\theta_q}$) are selected and trained, then q RTs generate q outputs ($\hat{Y}_1$,…, $\hat{Y}_q$). The values of the q outputs are then averaged to obtain the final output. The construction of RF is illustrated in **Fig. S4**.

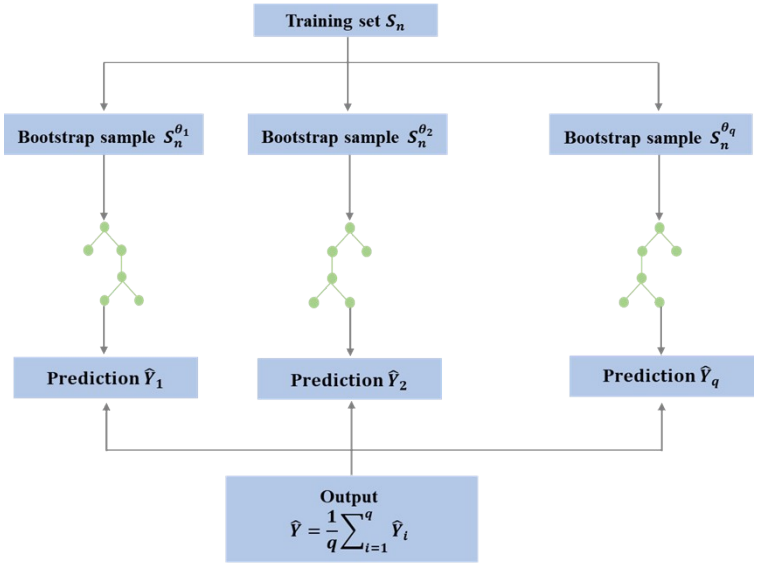

Fig. S4 Construction of an RF.

The BO process with five-fold CV:

Before model training, the dataset is randomly split into a training set containing 80% of data and a test set containing 20% of data. The training set was further split into five distinct subsets, with each subset serving as the validation subset in a different fold. In each iteration, the BO algorithm selected a new set of hyperparameters based on the prior knowledge. The ML model was trained based on four training subsets;

the loss was then computed on the corresponding validation subset to assess the model performance. After completing a 5-fold CV, the five sets of obtained loss values were averaged to get the final CV loss for that iteration. The goal of BO is to find the optimal combination of hyperparameters with minimum CV loss. The illustration diagram of ML model dataset partitioning used in this work is given in **Fig. S5**.

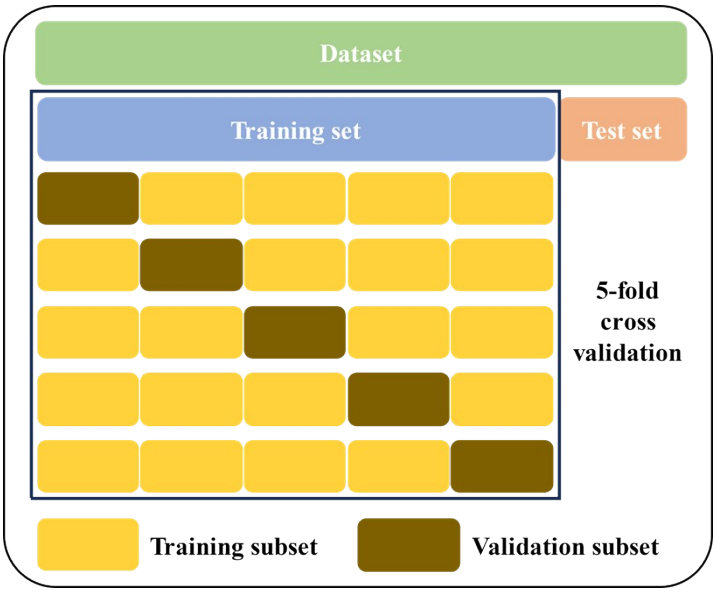

Fig. S5 The illustration diagram of five-fold cross-validation.

Table S1 Performance statistics of different models for SnSe SL prediction.

Performance parameters	R^2		RMSE		MAE	
	training	test	training	test	training	test
GPR	0.997	0.996	0.403	0.516	0.256	0.296
SVR	0.874	0.894	2.598	2.699	1.995	1.977
RT	0.971	0.943	1.251	1.981	0.935	1.4
RF	0.967	0.964	1.328	1.561	0.998	1.304

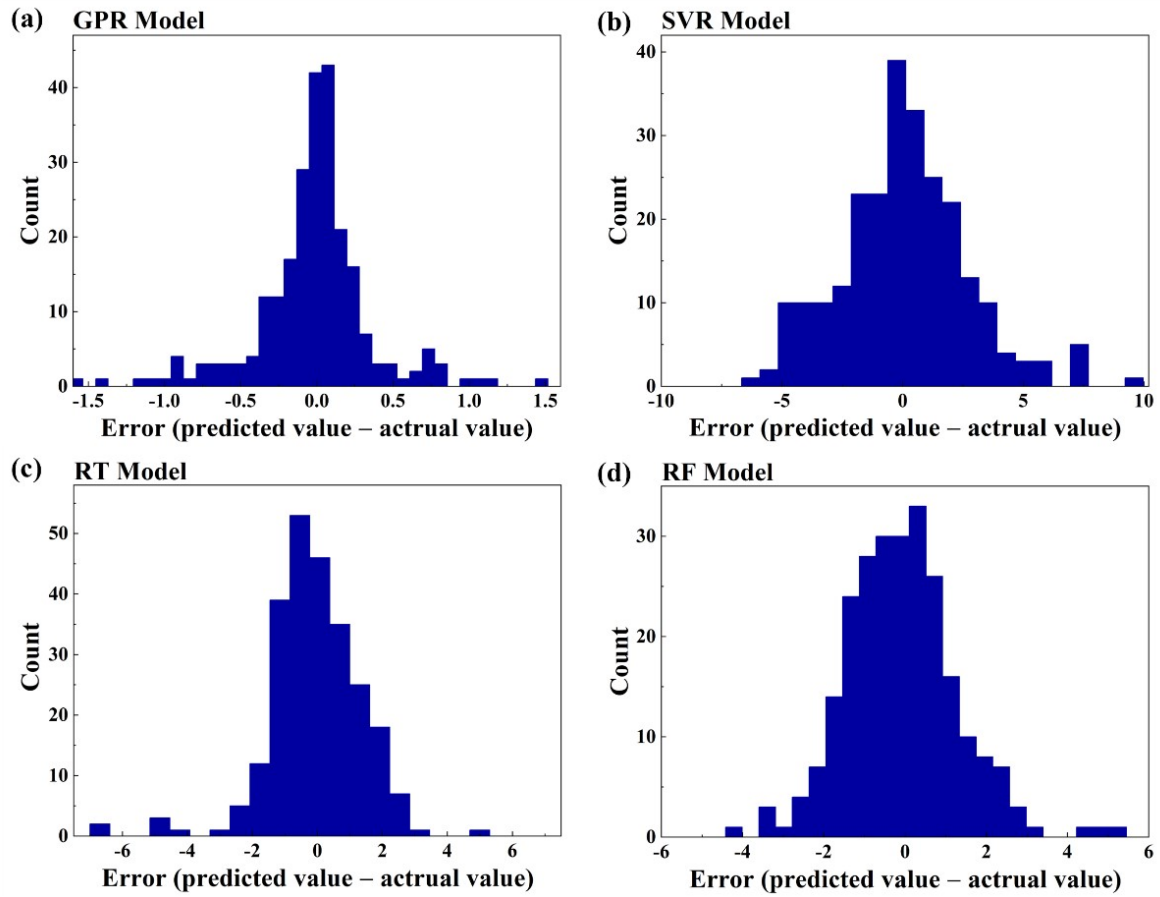

Fig. S6 The estimated error distributions between the predicted and actual SnSe NP SLs based on different ML models: (a) GPR model; (b) SVR model; (c) RT model and (d) RF model.

Characterization equipment

The surface morphology of the grown SnSe NPs was characterized with a SEM (FEI Verios 460 L). The chemical element composition of the grown SnSe NPs was analysed with micro-Raman spectroscopy (WITec alpha 300RA+) and EDX spectroscopy (Oxford Instruments X-Max 80 EDX system). The crystallographic information with atomic resolution of the grown SnSe NPs was confirmed with a HRTEM (FEI Titan G2 80-200 TEM/STEM) equipped with SAED, while a dual beam FIB-SEM system (FEI Helios NanoLab G3 CX) was used to prepare TEM specimens.