Data-efficient Materials Informatics and Inverse Design for Solid state Dewetting for Plasmonic applications

Shafayet Yaseen[1,2], Damjan Vengust[1], Vasyl Shvalya[1,2], Uroš Cvelbar[1,2]
[1]Jozef Stefan Institute, Ljubljana, Slovenia
[2]Jozef Stefan International Postgraduate School, Ljubljana, Slovenia
Corresponding author's email: shafayet.yaseen@gmail.com

Solid state thermal dewetting of thin metallic films provides a scalable physical pathway for self assembling nanoplasmonic particle arrays. However, exploring the multivariable fabrication space-ranging from initial film thickness, annealing temperature and complex substrate thermal dissipations-typically needs labor intensive trial and error mapping. To establish a data efficient predictive baseline without relying on big data deep learning architectures, we built and evaluated regularized, low complexity surrogate models on an experimentally obtained dataset of 32 silver thin film dewetting conditions across flat and 3D vertical nanostructured substrates (graphene and alumina nanowalls) [1].

Using Leave One Out cross validation with z-score standardization, predictive performance varied significantly across morphological targets. Mean equivalent particle diameter was redicted with high accuracy by smooth, continuous functions like Gaussian process regression with RBF kernel, and ridge regression, achieving R^2 values of 0.803 and 0.807, and MAE of 20.92 and 21.65 nm respectively, indicating global mass conservation and continuous growth kinetics. On the contrary, mean interparticle spacing and area fraction showed lower predictability: a simple ANN attained an R^2= 0.496, MAE= 23.71 nm, and ridge attained an R^2= 0.392 for the former, while a random forest yielded R^2= 0.510 and ridge attained R^2= 0.229 for the latter. This could potentially be a result of scalar averaging of interparticle spacing, that obscures the underlying substrate architecture by unifying two distinct topological population: local intrawall particles and interwall particles.

To address these limitations while maintaining physical grounding, the ongoing work adapts a two stage framework. First, we decouple the interparticle spacing target into local intrawall and interwall spatial distribution descriptors along with coupled UV-Vis spectral descriptors to eliminate scalar averaging error. Second, we carry out a pre specified 4 model feature ablation study under LOOCV: model A- baseline original 5 parameters, B- baseline + raw wall dimensions h, w, l and structural indicator, C- baseline + dimensionless physical descriptors including substrate aspect ratio and 3D surface area factor, and model D- combined baseline raw geometry and engineered descriptors. This benchmark aims to determine whether engineered dimensionless ratios generalize under data scarcity. Finally, validated forward models (extra trees, random forest, GPR, XGBoost) are coupled with genetic algorithms and Bayesian optimization for potential inverse recipe search with SHAP interpretability.

Acknowledgements
This work was supported by the Department of Gaseous Electronics (F6), Jožef Stefan Institute, and the Erasmus Mundus Joint Master programme GREENANO.

References
[1] S. Yaseen, "Localized Surface Plasmon Resonance Driven Laser Lithography on…", Master's Thesis, Jožef Stefan International Postgraduate School, Ljubljana, Slovenia (2026). COBISS.SI-ID 291882755.