Residual stress in sputtered and evaporated thin films: a ML approach

Tong Su

School of Engineering, Brown University, Providence RI 02912

GitHub(slides shows): https://github.com/vaccumwarrior2077/thinfilm_stress_ML.git

GitHub(for now, move to school email account): https://github.com/tongsu-brown/DATA1030_project.git

Introduction

Residual stress in thin films is a ubiquitous problem that affects their properties and performance. Examples include delamination and buckling [1], deformation of MEMS devices [2] and failures in interconnects [3]. Many studies have shown how stress depends on the material, growth conditions and microstructural evolution [4-6]. In previous work, we have developed explicit physical models and mathematical solutions for residual stress evolution in single element metal, binary solid solution and nitrides. But this method is highly constrained by physics mechanisms and assumptions. The testing error is very high when we split the data into training and testing sets. Therefore, we are seeking a way to loosen these constraints and give model more flexibility so that we can have a better understanding and prediction on the unknown processing conditions and unknown deposited materials.

To start, we rearrange the thin film stress data into the thin film residual stress database and apply a pure machine learning framework on the data. This approach can give us some understanding on how the data-driven method can fit the residual stress data and how in the future we can combine it with the previous built physical residual stress model to achieve the better interpretation and prediction.

EDA

For simplicity, we only give three major EDA interpretations for the thin film stress dataset which can cover most of the features. Figure 1 gives a basic picture of how the stress-thickness data looks like. For each material under certain processing conditions, the stress-thickness grows with the thickness of film. It gets higher with higher positive value (tensile stress) or get smaller with larger negative value (compressive stress). And the growing trend is different if the materials or the processing condition change. This interpretation shows that the stress-thickness is highly dependent on the different deposition materials and the processing conditions.

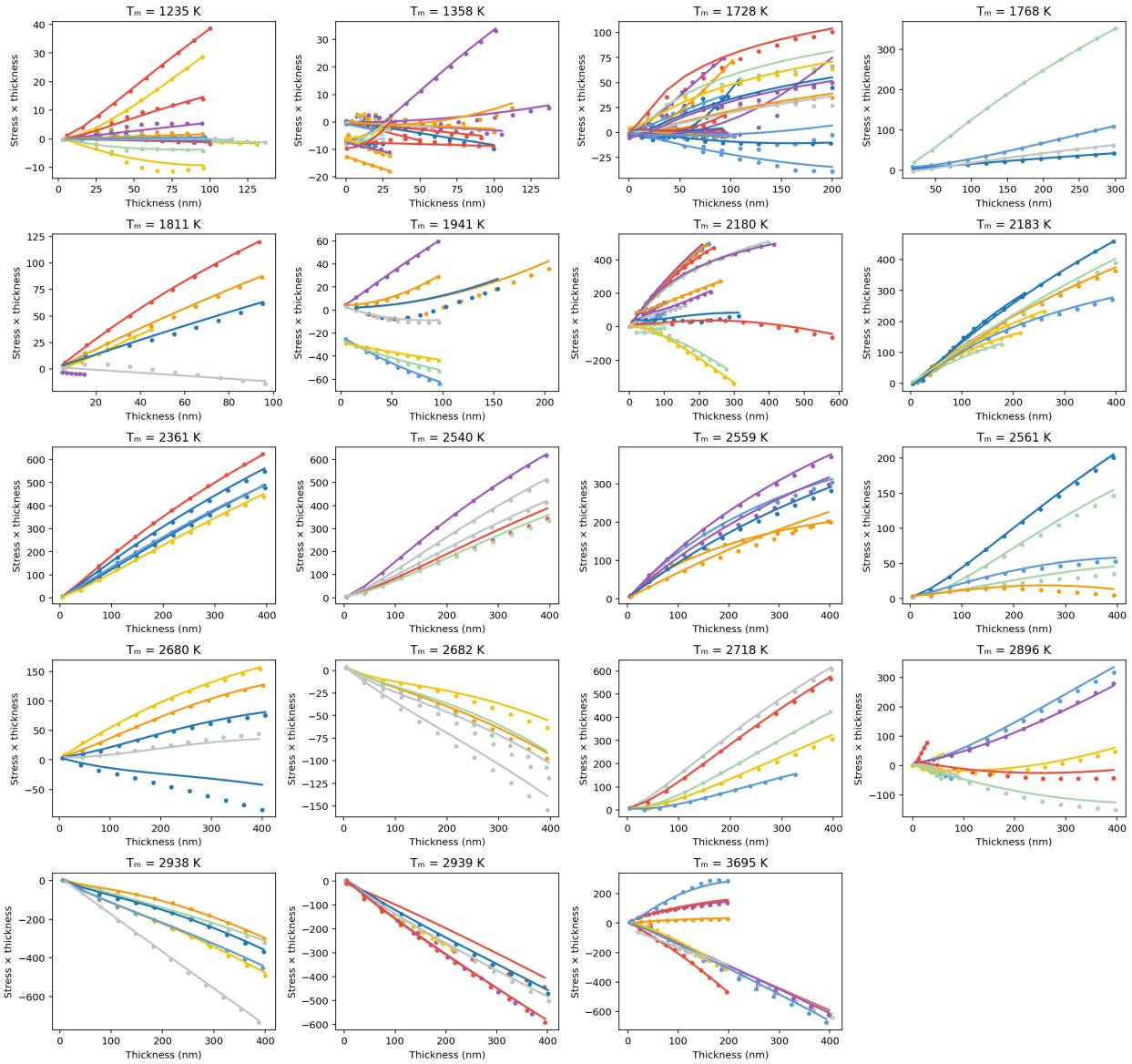

Figure 1, stress thickness raw data (dots) and analytical predictions (solid lines) for 19 different materials.

Figure 2 gives the number of the different crystal structures for the different kinds of alloy system. Crystal structure can be A15, BCC, FCC or HCP and alloy system is single element (just one element) or solid solution (two or more elements that can form complete solid solution). It indicates that the crystal structure may have some impact on stress-thickness. But from the stress mechanism speaking, the crystal structure should not affect thin film stress evolution. It can be a good point to test it in the importance evaluation to see whether this data pattern can against universal physics instinct.

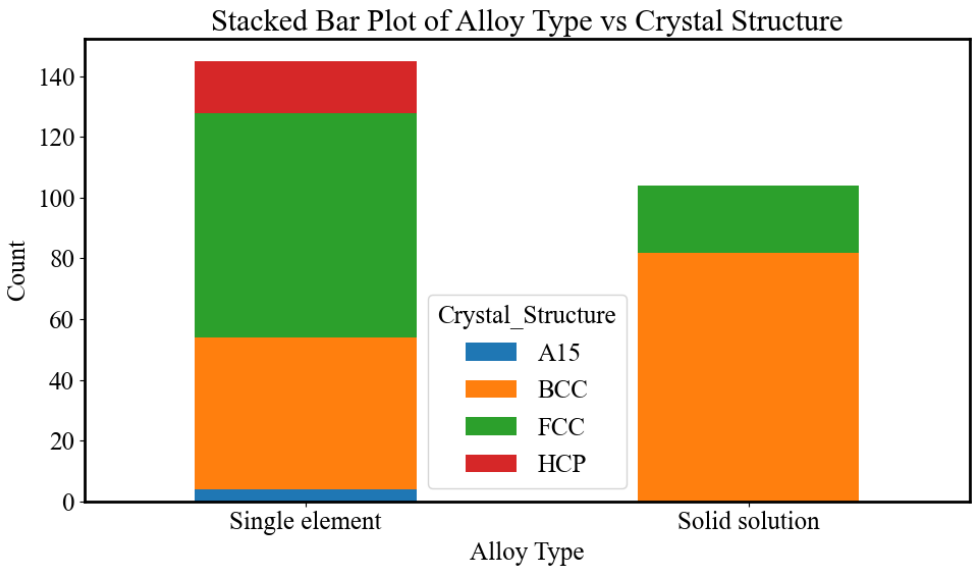

Figure 2, counts of the crystal structure for single element thin film and solid solution

Figure 3 represents the different kinds of substate we can have in the stress thickness dataset. The most common one is silicon with oxide layer. Some thin films were deposited on the metal layer or ionic layers. Same as the crystal structure, the physics instinct tell that this condition should have no impact on the thin film stress. But this parameter is found to be too complex as many datasets comes from the intermetallic alloys system, which is not included in the current testing yet. So, for this project, this feature is excluded from the final model tuning.

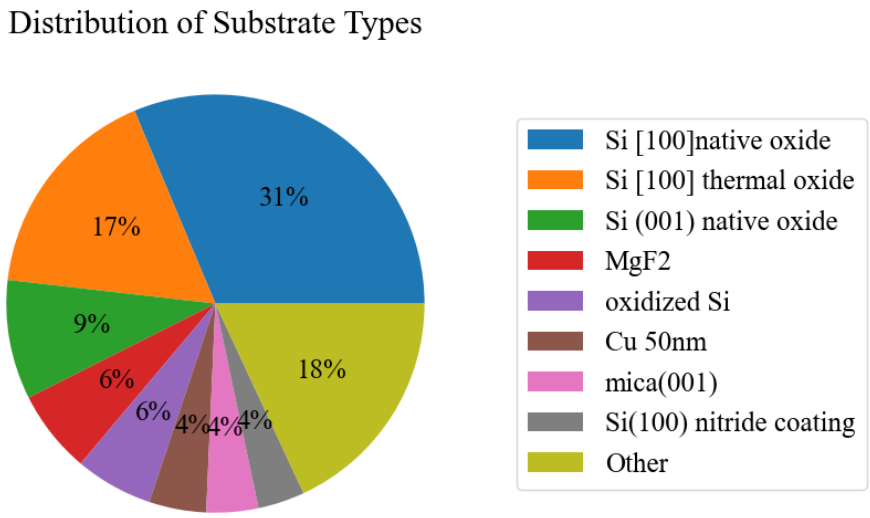

Figure 3, portion of the substrate type

Methods

In this project we treat the problem as supervised regression: using process conditions and basic material descriptors to predict the stress–thickness at a given thickness. We use a nested cross-validation scheme with inner and outer two loops. In the outer loop we perform ten randomized 80/20 train–test splits using a grouped shuffle split. For the first of the analysis, the dataset is grouped by material (on Tm_K, the melting point), so all curves from the same material stay together in either train or test, and the model is forced to predict the stress evolution for materials that it has never seen during training. We also repeat the same pipeline grouping by growth rate R to mimic predicting a new process window inside a material that the model already knows. Inside each outer train set, a 4-fold K-Fold cross-validation is applied to tune hyperparameters. The input features are: numeric features R (growth rate), P (pressure), T (substrate temperature), thickness and melting temperature Tm_K, and categorical features describing which lab or literature group the data came from, the alloy type and the crystal structure. Substrate type is excluded for the reasons discussed in the EDA section. All models share a single scikit-learn pipeline: a Column Transformer standardizes the numeric features using StandardScaler and one-hot encodes all categorical features with OneHotEncoder, and the transformed design matrix is passed to the chosen regressor. The missing values in categorical features care handled as the new class(no missing values in continuous features). This guarantees that every algorithm sees identically preprocessed inputs and prevents data leakage between training and testing folds.

We compare four regression algorithms that are covered in class. Ordinary least squares linear regression for Elastic Net as simple linear baselines. Random Forest, Gradient Boosting, Support Vector Machine and K nearest Neighbors are non-linear models that can capture strong interactions between process variables and non-linear dependence on thickness. During the inner cross-validation, root mean squared error (RMSE) is optimized because it has the same units as stress-thickness and penalizes large errors more strongly than mean absolute error; for interpretation we also report the coefficient of determination R^2 on both the inner and outer folds.

For Elastic Net, tuning parameters are regularization strength α (1e-4 - 1e3) and l1 degree (0.1 - 0.9). For Random Forest, tuning parameters are maximum tree depth (None, 2 - 20) and the fraction of features considered at each split (‘sqrt’, ’log2’, None). For Gradient Boosting, we tune the number of boosting stages (200 - 2000) and the learning rate (1e-3 – 0.5). For Support Vector, tuning parameters are different kernel method (‘rbf’, ‘linear’, ‘ploy’) and the C parameter (0.1 - 1000). For K Neighbors, we tune the number of neighbors (2 - 7) and the method to determine the weight ('uniform', 'distance').

As a very simple baseline we also fit a Dummy Regressor that always predicts the mean or median of the training stress–thickness which defines the baseline for RMSE/R^2 that it can compared against in the Results section. The spread of outer-loop scores across the ten group-shuffled splits provides an estimate of the uncertainty coming from different train–test partitions. To separate this from algorithmic stochasticity, we refit the Random Forest several times on the same split with different random seeds and compare the variance of the scores across seeds to the variance across outer splits.

Results

As a sanity check, the constant baseline model (Dummy Regressor) that always predicts the baseline gives $R^2 \approx 0$ by construction and a relatively large RMSE, so any useful model should do noticeably better than this. All four ML models clearly beat the baseline. The plain linear regression already recover a fair amount of the variance in stress-thickness except when splitting the dataset on materials (figure 6), but they tend to underfit materials with strong non-linear curvature in their stress–thickness curves. The two tree-based models perform best overall. As figure 4 shows, Gradient Boosting model reaches a best-fold test R^2 of 0.952, with an outer-loop out-of-fold (OOF) R^2 of 0.816, while the Random Forest reaches an even higher best-fold R^2 of 0.984 but a slightly lower mean OOF R^2 of 0.761. But the best test fold is just the random lucky one and cannot represent the general trend. So, in figure 5 we shows the average test R^2 over the random states with OOF. In this setting, the two best models have almost the same mean and OOF R^2. But SVR and KNN witness a much lower mean test value, which indicates that these two models are not stable when changing the grouping setting (some case the prediction is rather low accuracy). Overall, this behavior suggests that Random Forest and Gradient Bosting have both good effect on this dataset with growth rate R grouping method.

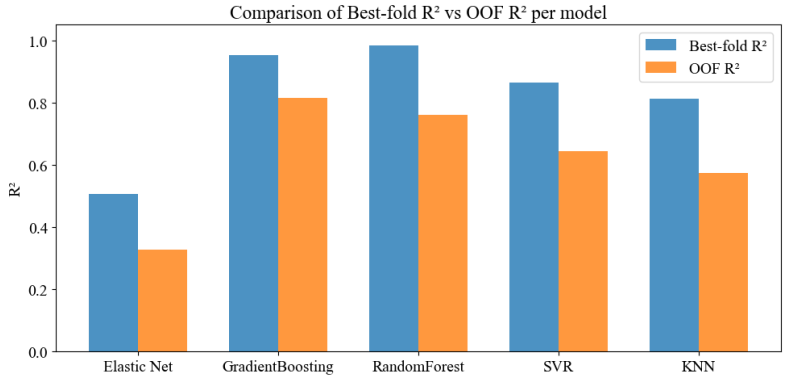

Figure 4, R^2 value of best test fold and OOF for data grouping by growth rate R

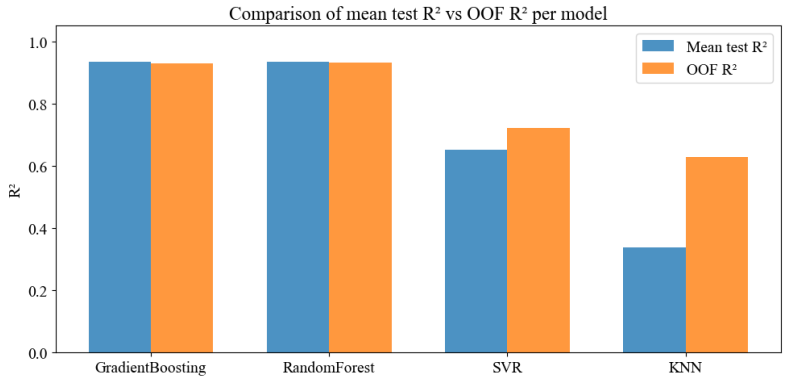

Figure 5, R^2 value of average test fold and OOF for data grouping by growth rate R

For the main evaluation I group by material (figure 4) in the outer split, so the model never sees any curve from a test material during training. The predicted vs. true curves grouped by material show that the Gradient Boosting model usually reproduces both the sign (tension vs. compression) and the overall growth trend of the stress–thickness evolution for unseen materials. Most of the larger discrepancies appear at very small thickness, where the data are noisy and the early Volmer-Weber coalescence stage is difficult to capture even with physics-based models. When I instead group by growth rate R (figure 5), the task becomes "predict a new process window" inside materials that the model has already seen. As expected, the scores drop somewhat and the uncertainty between different splits becomes larger, but the model still recovers the main trend of stress-thickness versus thickness. If I restrict the analysis to data from a single experimental group (only one lab or literature source, as figure 6 shows) the heterogeneity is reduced and both tree-based models give visibly tighter fits, which is consistent with the strong "group" effect that appears in the feature importance analysis. Also, in the scatter plot (figure 9), the OOF prediction is compared with the raw data grouping by growth rate and only from one lab source. The prediction is almost great and there are only several lines have totally out of range predictions.

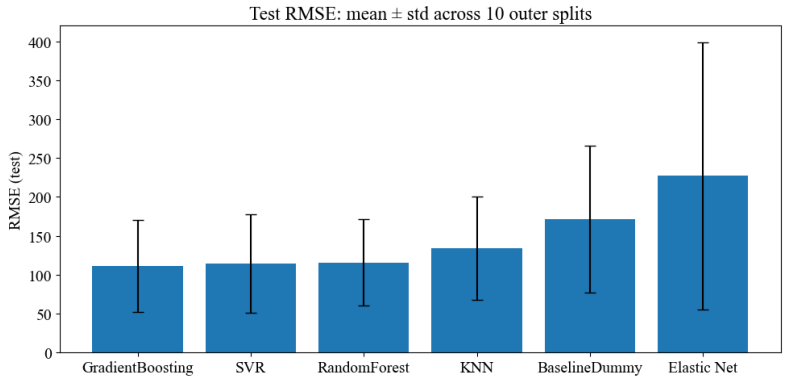

Figure 6, RMSE for different models on data grouping by materials (Tm_K)

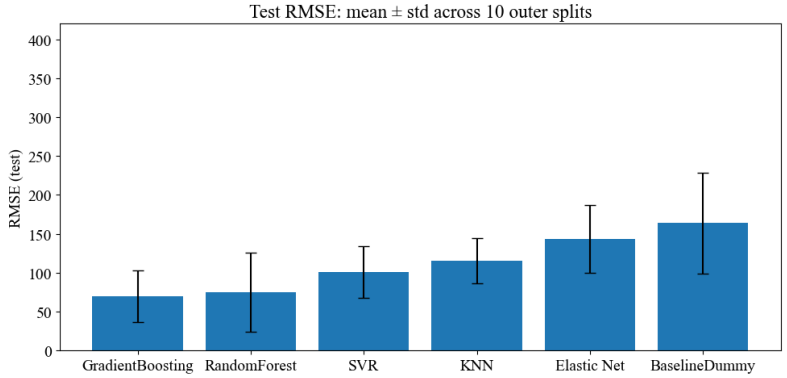

Figure 7, RMSE for different models on data grouping by growth rate R

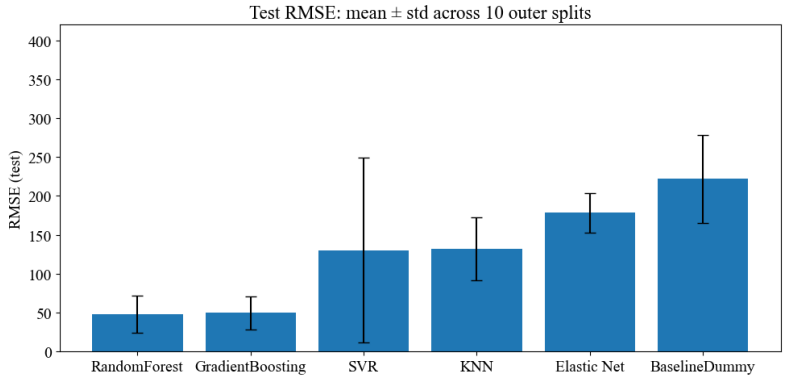

Figure 8, RMSE for different models on data grouping by growth rate R on only experiments performed by one lab condition

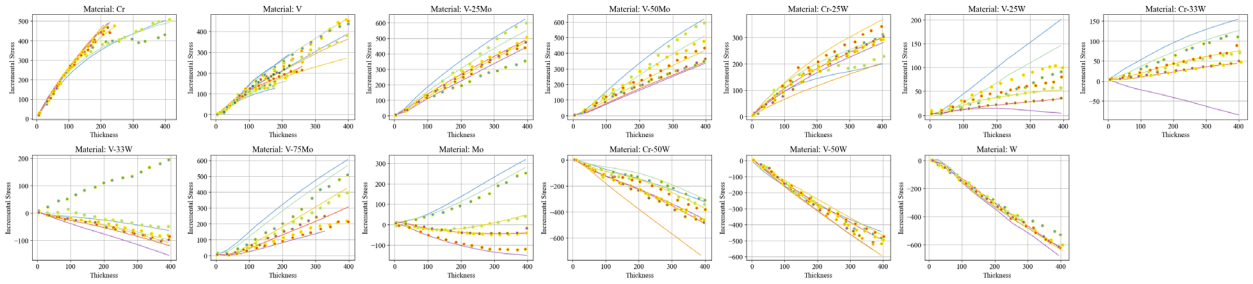

Figure 9, OOF scatter plot for the data grouping by R for only one lab source, the line are raw data from experiments and dots are OOF predictions

Global feature(figure 10) importance is computed for model that is grouped by growth rate R on only one lab source Su by permutation importance on the held-out data, shuffling one feature at a time and measuring the increase in RMSE. Tm_K, thickness, growth rate R and alloy type are consistently among the most important predictors. Melting temperature Tm_K has major weight: higher melting point materials tend to show more compressive steady-state slopes, which the model needs to learn. Thickness is the second important as the stress-thickness evolves with the thickness and they have one-by-one relation. Surprisingly the processing conditions (R, T, P) have much smaller importance when compared with the kind of material (Tm_K). It shows that model did not learn much from the processing conditions and heavily depend on the different kinds of the materials. In contrast, alloy crystal structure and alloy type are close to noise level once the main process variables are included. This is reasonable because the physical intuition was that crystal structure by itself should not dominate stress evolution, even though the raw counts in Figure 2 suggested some correlation.

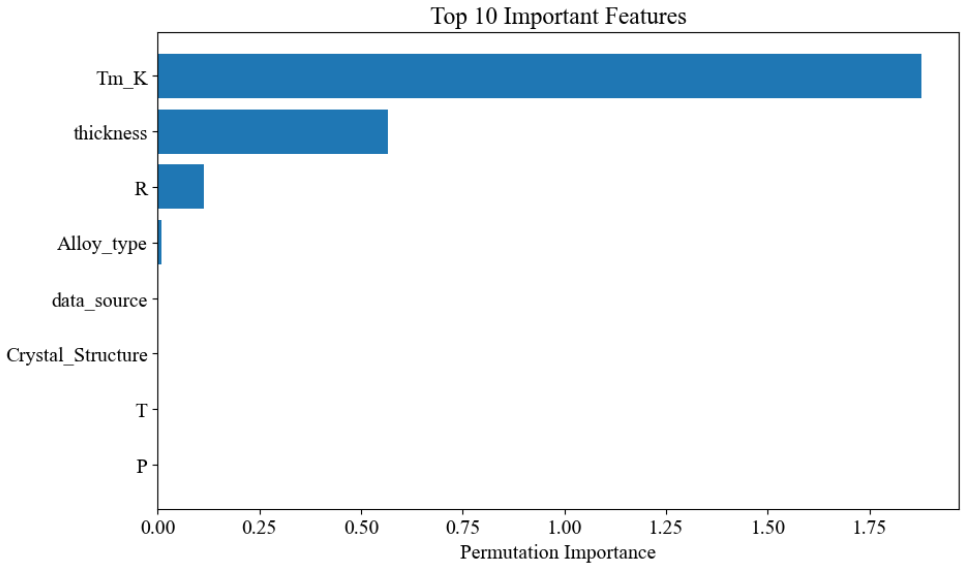

Figure 10, permutation importance for the data grouping by R for only one lab source

For local interpretation (Figure 11) I compute SHAP values for the Gradient Boosting model for a single representative sample and plot the top ten contributions. In this example the largest positive pushes on the predicted stress–thickness come from the melting temperature Tm_K and the film thickness, while the growth rate R slightly decreases the prediction. The categorical indicators (solid solution vs. single element, BCC structure, and the specific data source) make only small adjustments, and the contributions from T and P are essentially zero for this point. This local view is consistent with the global importance results: the model mainly uses physically meaningful process variables and basic material descriptors to set the overall level of stress–thickness, with the categorical features acting as minor corrections rather than dominating the prediction.

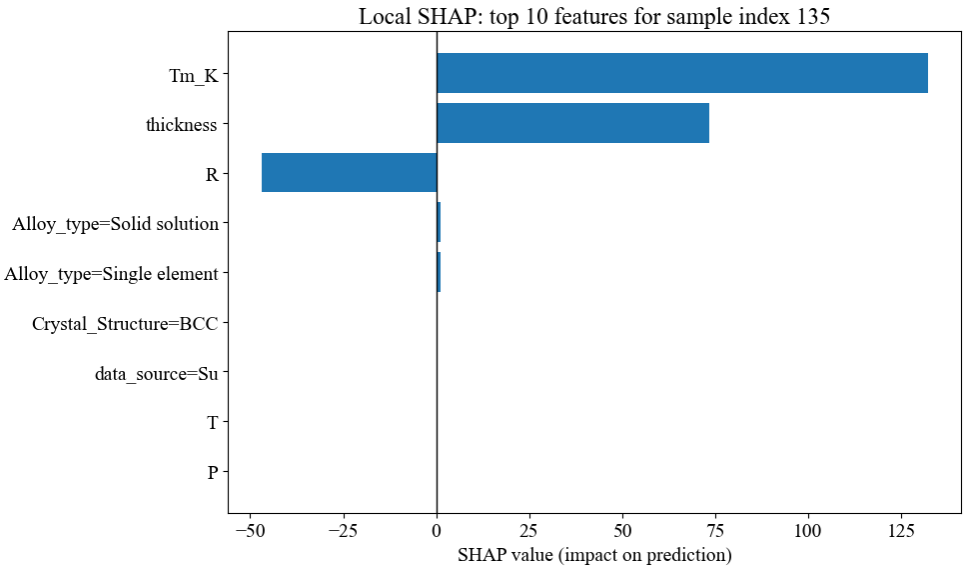

Figure 11, local importance for the data grouping by R for only one lab source

Outlook

The current model can only predict within certain experiments groups and predict the unknown process conditions (R, growth rate). From the suggestion of the importance matrix, the next step would be to add more physical constraint to the model to help it learn the relationship between stress-thickness and processing conditions. Like the energetic impacts (which describe the Vegard's law on the solid solution) can give better explanations on how the high atoms create compressive stress in the thin film.

The model only uses metallic data in the thin film stress dataset while the intermetallic thin films also have high value on the coating for energy, electrode and mechanical films. Therefore, if the ML model on the metallic film stress can be stabilized, we should extend the model to cover other kinds of the alloy system.

Steady state stress is another very important representation in thin film stress. When the thickness of film get larger (mostly higher than 200 nm), the slope of the stress-thickness curve becomes constant, and this slope is called steady state stress. Using steady state stress, we can analyze and predict the stress evolution in thicker films without knowing the exact deposition thickness. Figure 12 shows the steady state stress vs growth rate for V-W binary solid solution systems. In short, by adding more W content to the solution we can make stress more compressive and behave more like pure W. It would be another great direction to improve the ML model.

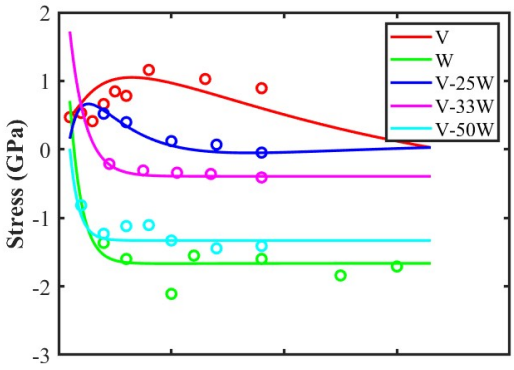

Figure 12, the steady state stress for V-W binary solid solution

Furthermore, we can add uncertainty estimation to the model. Like figure 13 shows, using uncertainty on the training model, we can have the respective uncertainty/error on each parameter and their get the uncertainty bands on the predictions. It can give better interpretation on our stress-thickness predictions: instead of just an absolute solution, we can give the upper and lower bands with certain criteria (75% or 95% confidence). It will be better instruction for the experiments tunning as we can have an expectation range for the stress level and gives more room for improvement on processing conditions.

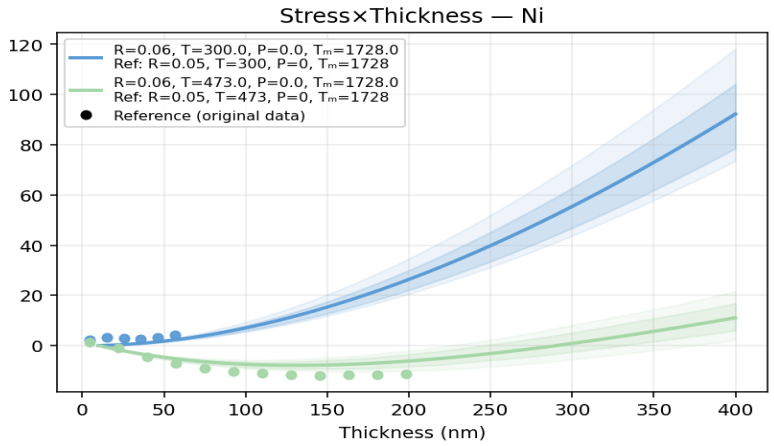

Figure 13, the prediction on unknown conditions (growth rate) for Ni with uncertainty bands.

Reference

[1] P. Waters, A.A. Volinsky, Stress and moisture effects on thin film buckling delamination, Exp. Mech. 47(1) (2007) 163-170.
[2] J. Matovic, Application of Ni electroplating techniques towards stress-free microelectromechanical system-based sensors and actuators, Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci. 220(11) (2006) 1645-1654.
[3] D. Vogel, E. Auerswald, J. Auersperg, P. Bayat, R.D. Rodriguez, D.R.T. Zahn, S. Rzepka, B. Michel, Stress analyses of high spatial resolution on TSV and BEoL structures, Microelectron. Reliab. 54(9-10) (2014) 1963-1968.
[4] R. Koch, The Intrinsic Stress of Polycrystalline and Epitaxial Thin Metal-Films, J. Phys.: Condens. Matter 6(45) (1994) 9519-9550.
[5] E. Chason, P.R. Guduru, Tutorial: Understanding residual stress in polycrystalline thin films through real-time measurements and physical models, J. Appl. Phys. 119(19) (2016) 191101.
[6] G. Abadias, E. Chason, J. Keckes, M. Sebastiani, G.B. Thompson, E. Barthel, G.L. Doll, C.E. Murray, C.H. Stoessel, L. Martinu, Review Article: Stress in thin films and coatings: Current status, challenges, and prospects, J. Vac. Sci. Technol. A 36(2) (2018) 020801.