Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Noise-Aware Machine Learning for Multimetal Surface Plasmon Resonance Parameter Estimation from Angular Reflectance Curves

Honey Thapa
Indian Institute of Technology Roorkee
Mayank Goswami
mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee https://orcid.org/0000-0001-8829-9980

Research Article

Keywords: optics, Photonics, AI

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11711593/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Noise-Aware Machine Learning for Multimetal Surface Plasmon Resonance Parameter Estimation from Angular Reflectance Curves

Honey Thapa[1] and Mayank Goswami[2]

[1]Department of Physics, IIT Roorkee
[2]MF School of AI and DS, IIT Roorkee

November 12, 2026

Abstract

Surface Plasmon Resonance (SPR) sensors are widely used for refractive index sensing due to their high sensitivity to changes in the surrounding medium. However, extracting multiple physical parameters from an SPR reflectance curve remains an inverse problem. This work investigates machine learning-assisted inverse modelling of SPR sensors in the Kretschmann configuration using reflectance spectra generated through the Transfer Matrix Method (TMM). Machine learning models are employed to estimate the analyte refractive index and metal thickness, while a metal classification model identifies the material.

A preliminary benchmark was performed to compare the machine learning approach with physics-based inverse modelling using TMM. On a noise-free held-out set of 60 samples, the machine learning model achieved a mean absolute error (MAE) of 0.003884 in refractive index and 0.21065 nm in metal thickness. The physics-based inverse model achieved errors of approximately 2.02×10^{-11} in refractive index and 2.88×10^{-8} nm in thickness. The machine learning approach provided a prediction time of 4.574 ms, compared with 44.038 ms for the physics-based inverse method, corresponding to an approximately 9.63-fold speed advantage in this benchmark.

These preliminary results demonstrate the trade-off between prediction accuracy and computational speed. Further evaluation under noisy conditions and across a broader parameter range is required to establish the robustness and generalisation capability of the proposed approach.

1 Introduction

Surface plasmon resonance (SPR) is widely used for refractive-index sensing because the coupling of p-polarized light to a surface plasmon mode produces a pronounced reflectance feature whose position and shape depend on the optical properties of the metal–dielectric system. In the Kretschmann configuration, a thin metal film is placed between a prism and the analyte. At a suitable incidence angle, energy couples into the surface plasmon mode and produces a dip in the reflected intensity. Changes in the analyte refractive index alter the angular reflectance response and can therefore be used for sensing.

In the forward problem, the optical response is calculated from known structural and material parameters. In the inverse problem studied here, the reflectance curve is given and the underlying

analyte refractive index and metal-film thickness are unknown. In this setting, using only the resonance angle may discard useful information contained in dip depth, width, asymmetry, and the broader curve shape. Full-curve parameter estimation is consequently a natural strategy, but it is also sensitive to the information content of the curve and to the assumptions used in the forward optical model.

Prior work has established that SPR reflectance curves can be fitted to recover layer properties, including thickness and dielectric parameters, and has examined noise and estimation uncertainty (Tilkens et al., 2000). Neural networks have also been used to estimate a liquid refractive index from simulated reflectance data for a prism-coupled SPR structure incorporating a chiral sculptured thin film, with robustness assessed under simulated noise (McAtee et al., 2019). Separately, deep learning has been applied to locate the SPR dip in angular-scanning measurements, with simulated noise added during training and validation against experimental measurements (Thadson et al., 2022). Thus, neither SPR curve inversion by itself nor noise-aware machine learning for SPR sensing is new in general.

A recent preprint by Runthala, Murai, and Arora uses TMM-generated data to build machine-learning surrogates for SPR performance metrics and combines them with explainability and optimization for sensor design (Runthala et al., 2026). That task maps known material and structural parameters to figures of merit and minimum reflectance and uses optimization to find sensor configurations. The present work instead considers an inverse mapping from a sampled angular reflectance curve to a metal identity, analyte refractive index, and metal thickness. This is a different prediction direction, although both projects use physics-generated SPR datasets and machine learning. Related studies on sensitivity optimization and graphene-assisted SPR highlight that calculated performance can depend strongly on the structural and material assumptions (Chen et al., 2022; dos Santos Almeida et al., 2022).

The purpose of this study is not to claim that a particular machine-learning algorithm is universally superior to physics-based fitting. Rather, we examine whether controlled noise augmentation improves the generalization of a multimetal curve-to-parameter pipeline under a spatially held-out test protocol. The work makes three bounded contributions:

1. It formulates a joint multimetal inverse-estimation pipeline that classifies Ag, Au, or Cu and estimates both analyte refractive index and metal-film thickness from the full sampled angular reflectance curve.

2. It compares clean-only training with two Gaussian-noise augmentation strategies using training-only spatial validation and a parameter-space block test, with test noise varied independently of model selection.

3. It quantifies aggregate and metal-specific errors and identifies the large deterioration in silver parameter recovery as a key limitation requiring further physical and identifiability analysis.

The present results are based entirely on synthetic data. In addition to evaluating the machine-learning pipeline, a preliminary quantitative comparison with nonlinear TMM inverse fitting is performed on a subset of the held-out test curves. This initial comparison considers noise-free curves with the true metal identity supplied to the TMM optimizer. The implications and limitations of this benchmark, together with the assumptions associated with the optical constants and noise model, are discussed below.

2 Optical model and synthetic dataset

2.1 Kretschmann geometry and TMM

The simulated stack is prism/metal/analyte. The incident wavelength is fixed at $\lambda = 650$ nm and p-polarized (TM) light is used. The prism refractive index is computed from the dispersion expression implemented in the MATLAB dataset-generation script. The complex dielectric functions of the metal films are taken at 650 nm from the optical-constant data of Babar and Weaver (Babar and Weaver, 2015); they are treated as fixed with respect to the simulated film thickness. This fixed-constant assumption simplifies dataset generation but does not represent the full thickness dependence that can occur in deposited metallic films. Experimental studies have reported variation in thin-film optical properties and uncertainty in their determination (Johnson and Christy, 1972; Lee et al., 2008; Tao et al., 2020), while deposition conditions and morphology are also important to plasmonic film quality (McPeak et al., 2015).

For a metal layer of thickness d_m, define the normalized transverse-wavevector/admittance terms used in the implementation as

$$q_p = \frac{\cos\theta}{n_p}, \tag{1}$$

$$q_a = \frac{\sqrt{n_a^2 - n_p^2 \sin^2\theta}}{n_a^2}, \tag{2}$$

$$q_m = \frac{\sqrt{\epsilon_m - n_p^2 \sin^2\theta}}{\epsilon_m}, \tag{3}$$

$$\gamma_m = \frac{2\pi d_m}{\lambda}\sqrt{\epsilon_m - n_p^2 \sin^2\theta}, \tag{4}$$

where n_p is the prism refractive index, n_a is the analyte refractive index, and ϵ_m is the complex relative permittivity of the metal. The characteristic matrix of the metal layer is

$$M = \begin{pmatrix} \cos\gamma_m & -i\sin\gamma_m/q_m \\ -iq_m\sin\gamma_m & \cos\gamma_m \end{pmatrix}. \tag{5}$$

The reflection coefficient calculated by the implemented TMM is

$$r_p(\theta) = \frac{(M_{11} + q_a M_{12})\, q_p - (M_{21} + q_a M_{22})}{(M_{11} + q_a M_{12})\, q_p + (M_{21} + q_a M_{22})}, \tag{6}$$

and the reflectance is $R(\theta) = |r_p(\theta)|^2$. The implementation uses consistent complex-valued calculations for the metal layer and analyte transverse wavevector. The complete MATLAB source should accompany any public release so that the sign convention and numerical implementation can be independently checked.

The metal dielectric functions used by the current dataset generator are summarized in Table 1. These are the values encoded in the project script and should be understood as model inputs, not as new measurements of the films.

Table 1: Complex relative permittivity values at 650 nm used in the current TMM dataset generator.

Metal	$\mathrm{Re}(\epsilon_m)$	$\mathrm{Im}(\epsilon_m)$
Ag	-19.458	0.43688
Au	-12.266	0.77889
Cu	-13.906	0.65890

2.2 Parameter grid and curve sampling

The analyte refractive index is varied from 1.33 to 1.40 in increments of 0.001, giving 71 values. Metal thickness is varied from 30 to 60 nm in increments of 1 nm, giving 31 values. Each of the three metals is simulated at every point on this grid, resulting in

$$N_{\text{curves}} = 3 \times 71 \times 31 = 6603 \tag{7}$$

synthetic clean curves. Each curve is sampled at incidence angles from 30° to 70° in increments of 0.01°, giving 4,001 reflectance values per curve. Therefore, each input vector is

$$\mathbf{x}_i = [R_i(30^\circ), R_i(30.01^\circ), \ldots, R_i(70^\circ)], \tag{8}$$

with three associated labels: metal class $m_i \in \{\mathrm{Ag}, \mathrm{Au}, \mathrm{Cu}\}$, analyte refractive index $n_{a,i}$, and metal thickness $d_{m,i}$. The inverse mapping is thus

$$\mathbf{x}_i \longmapsto (m_i,\, n_{a,i},\, d_{m,i}). \tag{9}$$

All curves are generated numerically, and no measured reflectance data are included in this dataset.

3 Machine-learning pipeline

3.1 Metal classification

The classifier is a support vector classifier (SVC) with a radial basis function kernel. To reduce the classifier input dimension, every fifth angular sample is retained, giving 801 features from the 4,001-point curve. The classifier pipeline standardizes the features and applies principal-component analysis with 40 components before the SVC. The selected saved configuration uses $C = 1$ and the "scale" setting for γ. The classifier is held fixed in the noise-augmentation comparison, so the reported changes in regression performance are not caused by retraining the metal classifier.

3.2 Metal-conditioned regression

After metal classification, a separate regression model is used for the predicted metal. Each metal-specific regressor uses an RBF-kernel support vector regression (SVR) model wrapped for the two continuous targets. The regression input uses the full 4,001-point curve. Input features are standardized, and the two regression targets are standardized during model fitting and transformed back to physical units for evaluation. Hyperparameters are selected using spatially grouped cross-validation. The SVR uses $\epsilon = 0.01$ in the standardized target space. The selected values of C and γ for the stored run are shown in Table 2.

Table 2: Selected per-metal SVR hyperparameters reported by the training run.

Metal	C	γ
Ag	10^4	2.4994×10^{-5}
Au	10^3	2.4994×10^{-5}
Cu	10^4	2.4994×10^{-5}

3.3 Spatial train/test split

The primary evaluation uses a spatial block split over the regular refractive-index/thickness grid. The held-out test block contains $n_a \in [1.360, 1.380]$ and $d_m \in [38, 52]$ nm. This gives 21 refractive-index values and 15 thickness values per metal, or 945 held-out curves in total. To reduce near-neighbor leakage, a buffer region extending 0.003 in refractive index and 3 nm in thickness around the test block is not used for training. The saved split contains 4,902 training curves, 945 test curves, and 756 buffer curves that are excluded from both. This design tests generalization to a withheld region of the parameter grid, not arbitrary extrapolation beyond the full parameter ranges.

During model selection, training-only validation uses five spatial folds and three repeats. The strategy-selection score is the mean normalized RMSE over both continuous outputs, with each prediction error normalized by its target's physical range before aggregation:

$$\mathrm{NRMSE} = \sqrt{\frac{1}{2N_v}\sum_{i=1}^{N_v}\left[\left(\frac{\hat{n}_{a,i} - n_{a,i}}{1.40 - 1.33}\right)^2 + \left(\frac{\hat{d}_{m,i} - d_{m,i}}{60 - 30}\right)^2\right]}. \tag{10}$$

Lower values indicate better validation performance. The test set is not used to choose the augmentation strategy.

3.4 Gaussian-noise augmentation

Three regression-training strategies are evaluated. The clean-only strategy retains the original training curves. For each augmented strategy, one additional noisy copy of every clean training curve is generated and paired with the same target values as its source curve. For the strategy called augment-to-0.005, the noise standard deviation for each noisy copy is drawn from $\{0.001, 0.0025, 0.005\}$. For augment-to-0.01, it is drawn from $\{0.001, 0.0025, 0.005, 0.01\}$. Noise is added independently at each angular sample:

$$R_{\mathrm{noisy}}(\theta_j) = R_{\mathrm{clean}}(\theta_j) + \eta_j, \qquad \eta_j \sim \mathcal{N}(0, \sigma^2). \tag{11}$$

No noisy version of a test curve is used in training. The fitted classifier and SVR hyperparameters are retained while the regressors are refitted to the clean-plus-noisy training data. Test performance is evaluated at $\sigma \in \{0, 0.001, 0.0025, 0.005, 0.01, 0.02, 0.03\}$ using ten independent test-noise realizations for each nonzero noise level; the clean case is deterministic. Noise is not clipped to $[0, 1]$. Consequently, at the largest test-noise level, a small fraction of noisy samples falls outside the physical reflectance interval; this is a property of the chosen stress-test protocol and a limitation when interpreting high-noise results.

3.5 Evaluation metrics

Continuous prediction performance is summarized using mean absolute error (MAE) and root mean squared error (RMSE) for analyte refractive index and film thickness. Classification performance is reported using accuracy and macro-F1. The MAE for refractive index is dimensionless and represents refractive-index units (RIU); thickness MAE is reported in nanometres. Reported standard deviations for noisy test cases quantify variability across the ten independent noise realizations, not uncertainty over different physical samples or different dataset-generation assumptions.

4 Results

4.1 Training-only validation selects augmentation up to 0.01

Table 3 and Fig. 1 show the training-only validation scores. Clean-only training gives a mean normalized RMSE of 0.01206 ± 0.00998. Augmentation up to 0.005 reduces this to 0.00829 ± 0.00555, while augmentation up to 0.01 gives 0.00584 ± 0.00284. The last strategy reduces the mean score by approximately 51.6% relative to clean-only training. The dispersion across validation conditions is non-negligible, especially for clean-only training, so the mean score should be interpreted together with the spread.

Table 3: Training-only spatial-validation normalized RMSE. Values are mean ± standard deviation across the validation folds, repeats, and validation-noise conditions.

Training strategy	Mean NRMSE	Relative reduction vs clean
Clean-only	0.01206 ± 0.00998	—
Augment-to-0.005	0.00829 ± 0.00555	31.2%
Augment-to-0.01	0.00584 ± 0.00284	51.6%

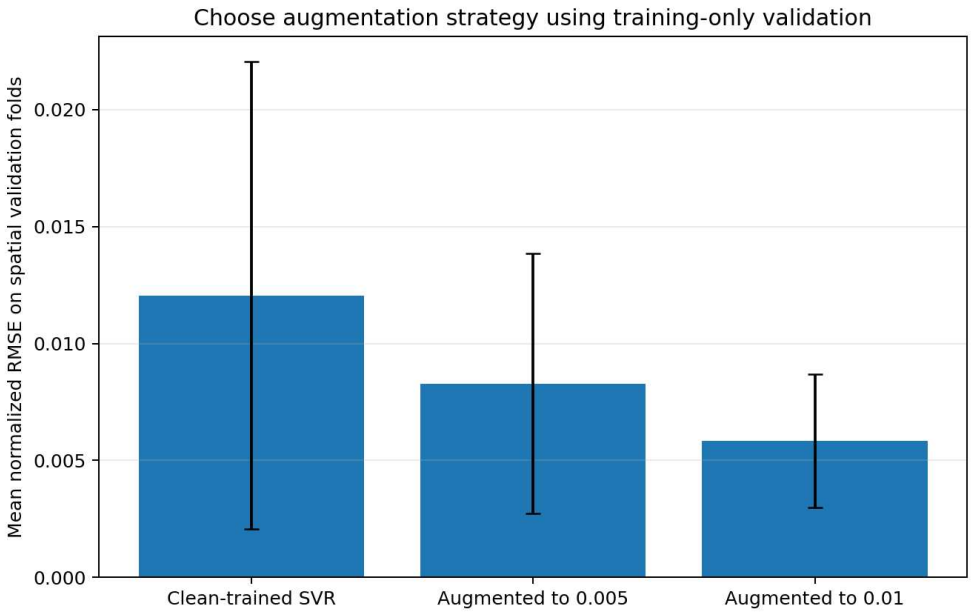


Figure 1: Selection of the noise-augmentation strategy using training-only spatial validation. Error bars show the standard deviation across validation conditions in the reported evaluation.

4.2 Classification of the metal

The SVC achieved 100% accuracy and macro-F1 on the 945-curve held-out block test set. The test set contains 315 curves for each of Ag, Au, and Cu, and no class errors were recorded for the tested noise levels. These results indicate that the three classes are separable within the present synthetic configuration and parameter ranges. They do not establish perfect metal identification for real measurements, under optical-constant mismatch, or outside the simulated range.

4.3 Aggregate regression errors versus test noise

Table 4 and Fig. 2 summarize refractive-index MAE. At zero test noise, the clean-only strategy is slightly better than augmentation up to 0.01. As test noise increases, the relative advantage of the broader augmentation strategy grows. At $\sigma = 0.01$, RI MAE decreases from 1.875×10^{-3} for clean-only training to 8.26×10^{-4} for augmentation up to 0.01, a reduction of approximately 56.0%. At $\sigma = 0.03$, the corresponding reduction is approximately 69.0%.

Table 4: Held-out block refractive-index MAE (RIU) averaged over ten test-noise realizations at each nonzero σ.

σ	Clean-only	Augment-to-0.005	Augment-to-0.01
0	4.069×10^{-4}	4.033×10^{-4}	4.329×10^{-4}
0.001	4.130×10^{-4}	4.087×10^{-4}	4.359×10^{-4}
0.0025	4.548×10^{-4}	4.395×10^{-4}	4.528×10^{-4}
0.005	7.130×10^{-4}	5.790×10^{-4}	5.243×10^{-4}
0.01	1.875×10^{-3}	1.225×10^{-3}	8.256×10^{-4}
0.02	6.335×10^{-3}	3.836×10^{-3}	2.080×10^{-3}
0.03	1.299×10^{-2}	7.763×10^{-3}	4.026×10^{-3}

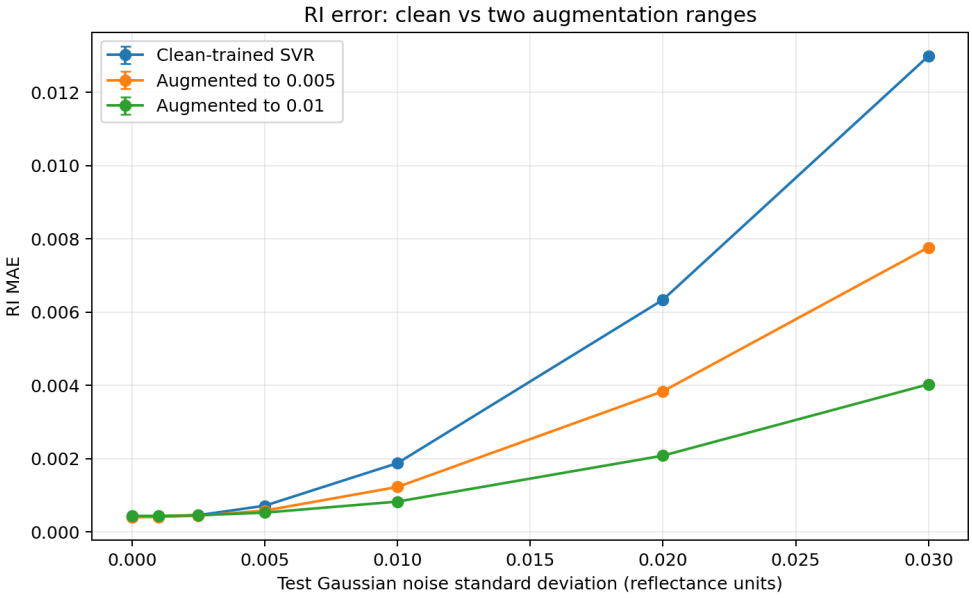


Figure 2: Held-out block refractive-index MAE as a function of test-noise standard deviation for the three training strategies. Error bars show variability across the independent test-noise realizations.

Thickness MAE follows the same pattern (Table 5, Fig. 3). At $\sigma = 0.01$, augmentation up to 0.01 reduces thickness MAE from 0.1547 nm to 0.0972 nm, a reduction of approximately 37.2%. At $\sigma = 0.03$, it decreases from 0.8151 nm to 0.3769 nm, a reduction of approximately 53.8%. The augmentation strategy has a modest clean-data cost: at $\sigma = 0$, thickness MAE is 0.0753 nm for augment-to-0.01, compared with 0.0689 nm for clean-only training.

Table 5: Held-out block metal-thickness MAE (nm) averaged over ten test-noise realizations at each nonzero σ.

σ	Clean-only	Augment-to-0.005	Augment-to-0.01
0	0.06895	0.07146	0.07530
0.001	0.06976	0.07195	0.07553
0.0025	0.07465	0.07461	0.07627
0.005	0.09154	0.08419	0.07962
0.01	0.1547	0.1186	0.09719
0.02	0.4157	0.2737	0.2038
0.03	0.8151	0.5121	0.3769

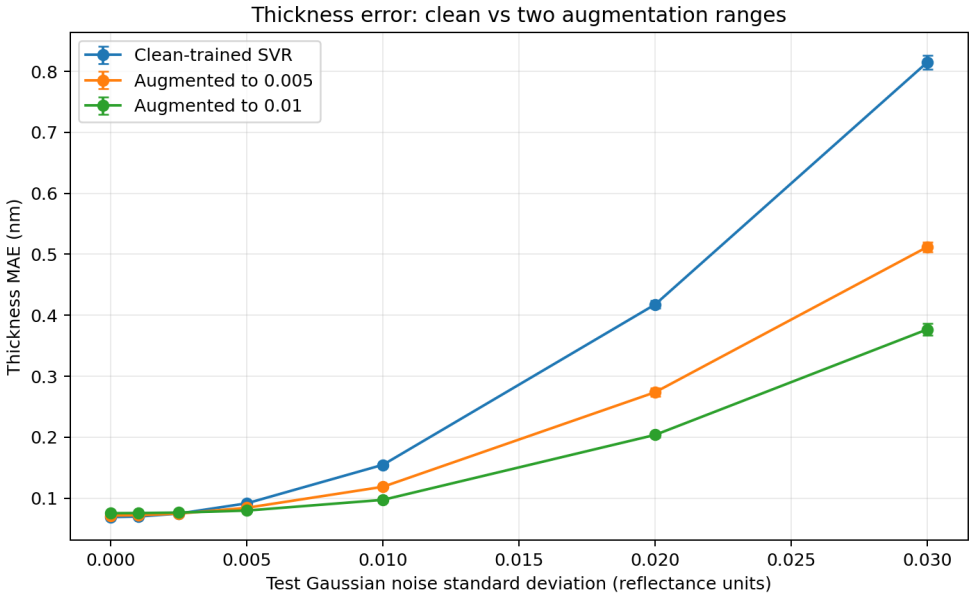


Figure 3: Held-out block metal-thickness MAE as a function of test-noise standard deviation. Wider noise augmentation improves robustness at moderate and high noise, while increasing error slightly on clean curves.

4.4 Differences between metals

The aggregate metrics conceal substantial differences between metals. Silver has the largest parameter-recovery errors, whereas gold has the smallest errors over much of the tested range. Figure 4 shows the RI error by metal, and Fig. 5 shows the thickness error by metal. For the augment-to-0.01 strategy at $\sigma = 0.03$, the per-metal thickness MAEs are approximately 0.825 nm for Ag, 0.126 nm for Au, and 0.180 nm for Cu. The corresponding RI MAEs are approximately 0.00903, 0.000915, and 0.00214 RIU.

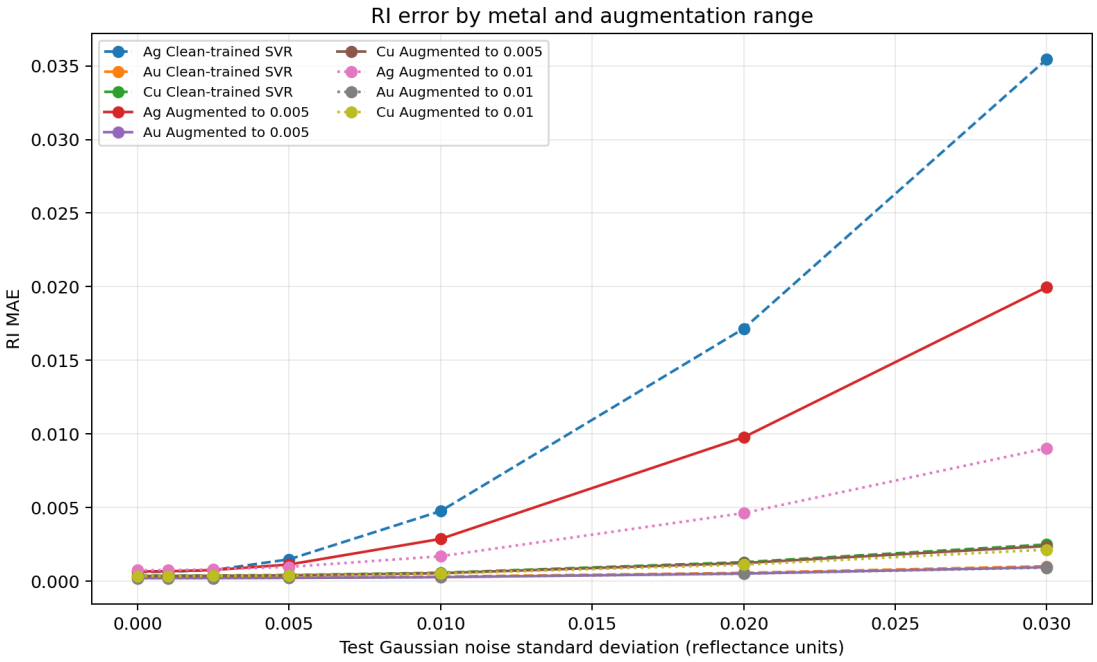


Figure 4: Held-out refractive-index MAE versus noise level, separated by metal and training strategy. Silver exhibits the strongest deterioration under noise in this simulation.

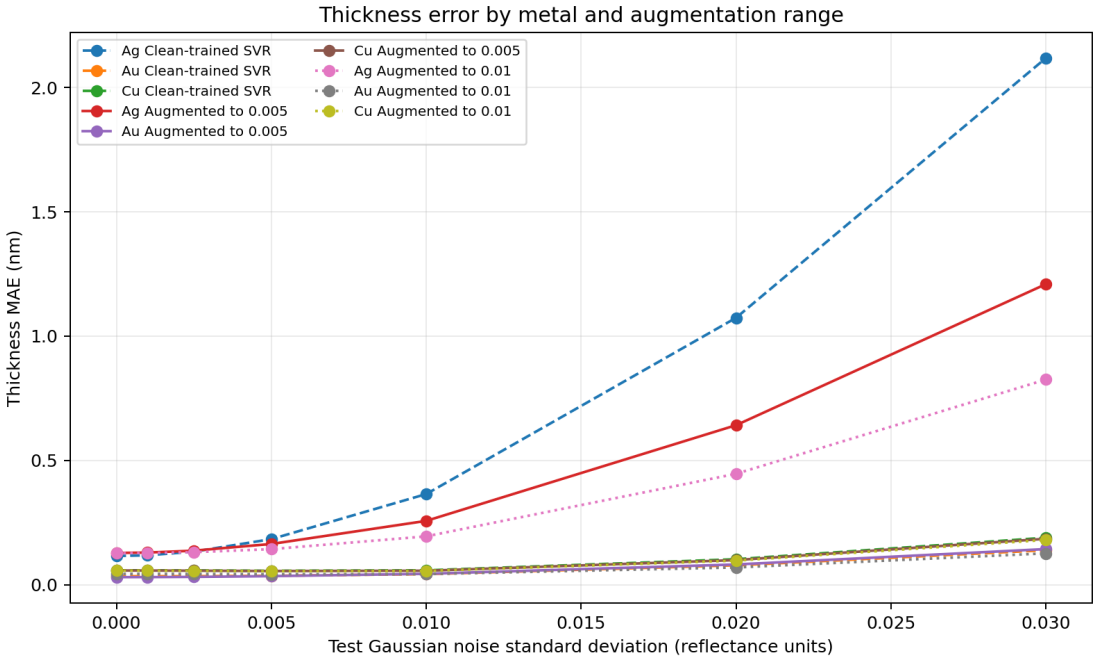


Figure 5: Held-out metal-thickness MAE versus noise level, separated by metal and training strategy.

These differences are empirical observations; their physical causes have not yet been isolated. Possible factors include differences in the reflectance-curve geometry, sensitivity of the curve to the two estimated parameters, and interactions between RI and thickness. A dedicated identifiability

and sensitivity analysis is required before assigning a causal explanation. In addition, the adopted constants are fixed with thickness, so this model cannot represent the thickness-dependent optical properties reported for some thin metal films (Lee et al., 2008; Tao et al., 2020).

4.5 Comparison with Inverse TMM Fitting

A preliminary benchmark was conducted to compare the proposed machine-learning pipeline with nonlinear inverse fitting using TMM. Both methods were evaluated on the same 60 noise-free curves ($\sigma = 0$) from the held-out test block.

The inverse TMM method estimated the analyte refractive index (n_a) and metal thickness (d_m) by minimizing the sum of squared differences between the observed and simulated reflectance:

$$L(n_a, d_m) = \sum_{j=1}^{L} \left[R_{\text{obs}}(\theta_j) - R_{\text{TMM}}(\theta_j; n_a, d_m)\right]^2 . \tag{12}$$

Bounded nonlinear least squares with five initial guesses was used. The true metal identity was supplied to the optimizer, so only the two continuous parameters were estimated.

Table 6: Comparison of ML inference and inverse TMM fitting on 60 noise-free held-out curves.

Method	RI MAE (RIU)	Thickness MAE (nm)	Time (ms)
ML (single-curve)	3.88×10^{-3}	0.2107	4.57
ML (batched)	3.88×10^{-3}	0.2107	0.62
TMM (known metal)	2.02×10^{-11}	2.88×10^{-8}	44.04

TMM achieved substantially lower parameter errors, while ML provided approximately 9.63 times faster single-curve inference. Batched ML further reduced the average time to 0.62 ms per curve. The TMM optimizer successfully converged for all 60 curves, and the ML classifier correctly identified all 60 metal identities.

The near-zero TMM errors are expected because the synthetic data and inverse fitting use the same forward optical model. Therefore, these results demonstrate parameter recovery under ideal model conditions rather than expected accuracy on experimental data.

5 Discussion

5.1 What the current results establish

Within the specified synthetic parameter grid and split, noise augmentation improves regression robustness under the independent additive Gaussian perturbations studied here. The model-selection results are consistent with the test results: augmentation up to 0.01 gives the lowest validation NRMSE and the lowest aggregate regression errors for the higher test-noise levels. Conversely, the clean-only strategy is slightly better when the test curves contain no added noise. This is a useful trade-off rather than evidence that one strategy dominates in all conditions.

The held-out parameter block is an important element of the evaluation. Because nearby RI/thickness combinations are withheld together, the test is less optimistic than an ordinary randomly sampled split over the same regular parameter grid. Still, it is an interpolation/generalization

test within the global parameter ranges, not an assessment of extrapolation to new wavelengths, new prism materials, new layer stacks, new optical constants, or experimental measurements.

5.2 Relation to prior inverse and ML studies

Classical SPR fitting already addresses the recovery of unknown layer properties from reflectance curves, including uncertainty under noise (Tilkens et al., 2000). ML has also been used to estimate RI from simulated SPR-related optical data and to improve angular-dip localization under simulated noise, with the latter validated experimentally (McAtee et al., 2019; Thadson et al., 2022). The 2026 Runthala et al. preprint addresses ML surrogate modeling and optimization for sensor design, where structural inputs predict performance outputs (Runthala et al., 2026). Accordingly, the potential contribution here is narrow: a controlled study of noise-aware ML for simultaneous metal identification and recovery of two continuous parameters from full angular curves, evaluated on held-out parameter regions and separated by metal. This positioning should be revisited after a broader literature search; the present reference set is not exhaustive.

5.3 Limitations

Several limitations materially restrict the conclusions.

1. **Synthetic-only evaluation.** The forward model generates all curves. No experimental measurements are used, so the reported MAEs do not establish real-instrument accuracy or precision.
2. **Fixed metal dielectric functions.** The optical constants are fixed across thickness, despite literature evidence that the properties of deposited thin films can depend on thickness and preparation (McPeak et al., 2015; Lee et al., 2008; Tao et al., 2020).
3. **Simplified noise.** Independent additive Gaussian noise is not a complete instrument-noise model. At high σ, some reflectance samples fall outside $[0, 1]$ because the curves are deliberately not clipped; such inputs are not physically realizable reflectance values.
4. **No completed physics-based baseline in the reported table.** A nonlinear least-squares TMM inverse fit is a natural baseline for this inverse problem, but no directly comparable, verified result is reported here. Therefore, this manuscript does not claim that ML is more accurate or more reliable than physics-based fitting.
5. **Unresolved identifiability.** A systematic analysis of near-degenerate (n_a, d_m) solutions and sensitivity of the reflectance curve has not yet been completed. Large silver errors should not be attributed to one physical mechanism without this analysis.
6. **Optical constants and sign conventions require reproducible release.** The full generator, dielectric data, complex square-root handling, and convention for positive loss should be published and checked independently.

5.4 Next Steps

The preliminary inverse-TMM benchmark provides an initial comparison of parameter-estimation accuracy and computational cost. Further evaluation is required to establish the relative merits of both approaches.

First, the comparison should be extended to noisy held-out curves using identical test samples and noise realizations for both methods, across $\sigma \in \{0, 0.005, 0.01, 0.02\}$. A complete TMM baseline should also test all three metal hypotheses (Ag, Au, and Cu) for each curve, selecting the material with the lowest final reflectance residual. This would enable a fairer comparison with the proposed ML classifier and metal-conditioned regressors.

Further analysis should consider parameter errors, metal-identification accuracy, convergence failures, sensitivity to initialization, and runtime. An identifiability study and tests involving angular offsets, correlated noise, and optical-constant mismatch would also help assess model robustness. Finally, validation using experimentally measured reflectance curves is necessary to establish practical applicability.

6 Conclusion

This work evaluated a noise-aware machine-learning pipeline for estimating analyte refractive index and metal-film thickness from SPR reflectance curves for Ag, Au, and Cu. Training-only validation showed that noise augmentation up to $\sigma = 0.01$ reduced the mean normalized RMSE by approximately 51.6% compared with clean-only training. At test noise levels of $\sigma = 0.01$ and 0.03, the pipeline achieved RI MAEs of 8.26×10^{-4} and 0.00403 RIU, respectively, with corresponding thickness MAEs of 0.0972 and 0.377 nm.

A preliminary comparison with inverse TMM fitting on 60 noise-free held-out curves showed that TMM achieved substantially lower parameter errors, with RI and thickness MAEs of 2.02×10^{-11} RIU and 2.88×10^{-8} nm. The ML pipeline was approximately 9.63 times faster, requiring 4.57 ms per curve compared with 44.04 ms for TMM fitting. However, this benchmark used the known metal identity and did not include noisy curves or automatic metal identification by TMM.

Overall, noise augmentation improved the robustness of the ML pipeline within the tested synthetic conditions, while inverse TMM provided greater accuracy at a higher computational cost. Further end-to-end benchmarking and experimental validation are necessary to establish the practical applicability of the proposed approach.

Data and code availability

The synthetic dataset is generated by the MATLAB TMM code described above. The project files include the multimetal generator, training and augmentation scripts, split-generation utilities, configuration files, and summary CSVs. At the time of this draft, no permanent public repository URL or DOI is specified. Before public release, the exact dataset version used for the reported tables, test split indices, saved model, and software versions should be deposited in a permanent public repository.

Author contributions

The author field and contribution statement should be revised to reflect the actual contributions of all contributors and any required institutional or supervisory guidance before public posting.

Competing interests

The author should confirm whether any competing interests exist before submission.

References

Shaista Babar and J. H. Weaver. Optical constants of Cu, Ag, and Au revisited. *Applied Optics*, 54(3):477–481, 2015. doi: 10.1364/AO.54.000477.

Shujing Chen, Zhining Lin, Gaiyan Bai, and Chengyou Lin. Comparative study on sensitivity enhancement of a graphene based nearly guided-wave surface plasmon resonance biosensor optimized using genetic algorithm in the visible region. *Optical and Quantum Electronics*, 54(3):199, 2022. doi: 10.1007/s11082-022-03584-0.

Aline dos Santos Almeida, Dario A. Bahamon, Nuno M. R. Peres, and Christiano J. S. de Matos. A critical analysis on the sensitivity enhancement of surface plasmon resonance sensors with graphene. *Nanomaterials*, 12(15):2562, 2022. doi: 10.3390/nano12152562.

P. B. Johnson and R. W. Christy. Optical constants of the noble metals. *Physical Review B*, 6(12):4370–4379, 1972. doi: 10.1103/PhysRevB.6.4370.

Wook-Jae Lee, Jae-Eun Kim, Hae Yong Park, Suntak Park, Min su Kim, Jin Tae Kim, and Jung Jin Ju. Optical constants of evaporated gold films measured by surface plasmon resonance at telecommunication wavelengths. *Journal of Applied Physics*, 103(7):073713, 2008. doi: 10.1063/1.2902395.

Patrick D. McAtee, Satish T. S. Bukkapatnam, and Akhlesh Lakhtakia. Artificial neural network to estimate the refractive index of a liquid infiltrating a chiral sculptured thin film. *Journal of Nanophotonics*, 13(4):046006, 2019. doi: 10.1117/1.JNP.13.046006.

Kevin M. McPeak, Sriharsha V. Jayanti, Stephan J. P. Kress, Stefan Meyer, Stelio Iotti, Aurelio Rossinelli, and David J. Norris. Plasmonic films can easily be better: Rules and recipes. *ACS Photonics*, 2(3):326–333, 2015. doi: 10.1021/ph5004237.

R. Runthala, S. Murai, and P. Arora. Machine learning-assisted analysis and inverse design of prism-based surface plasmon resonance sensors, 2026. URL https://arxiv.org/abs/2608.17368. Preprint.

Longbiao Tao, Shuo Deng, Hongyun Gao, Haifei Lv, Xiaoyan Wen, and Min Li. Experimental investigation of the dielectric constants of thin noble metallic films using a surface plasmon resonance sensor. *Sensors*, 20(5):1505, 2020. doi: 10.3390/s20051505.

Kitsada Thadson, Suvicha Sasivimolkul, Phitsini Suvarnaphaet, Sarinporn Visitsattapongse, and Suejit Pechprasarn. Measurement precision enhancement of surface plasmon resonance based angular scanning detection using deep learning. *Scientific Reports*, 12:2052, 2022. doi: 10.1038/s41598-022-06065-2.

Bernard G. Tilkens, Yves F. Lion, and Yvon L. Renotte. Uncertainties in the values obtained by surface plasmon resonance. *Optical Engineering*, 39(2):363–373, 2000.